Using PDA consistency checks to increase the precision of profits and sales measurement in panels[#]

Marcel Fafchamps, *Oxford University*

David McKenzie, *World Bank*

Simon Quinn, *Oxford University*

Christopher Woodruff, *University of Warwick*

Abstract

Personalized Digital Assistants (PDAs) and other forms of hardware needed to collect survey data electronically have become more affordable and powerful in recent years, leading to their use in a number of surveys in developing countries. Simple use of these devices can offer the prospect of more timely data entry and greater accuracy in guiding respondents through skip patterns. Further benefits are possible through the use of more complex consistency checks. We use PDAs to measure sales and profits for microenterprises, which are notoriously noisy. Consistency checks in the cross-section compare sales and profits, while those in the panel query responses which result in large changes from one period to the next. Cross-sectional checks also served as a second prompt in the case of missing profits. These checks do succeed in reducing the standard deviation and in increasing the correlation of the observations for which corrections are made. However, we find that the vast majority of large changes in enterprise sales and profits are confirmed by firm owners as genuine, highlighting the volatility of income in this sector. As a result, the overall impact of these consistency checks on the full sample is rather limited, suggesting that while such checks are useful if computerized forms of data collection are being used, the consistency checks per se are not a strong reason for using computerized data collection in collecting firm profits and sales.

JEL codes: C81, O12, M41

Keywords: Microenterprises; Measurement; Profits; Panel; Survey Methods; Electronic Data Collection.

[#] The authors thank the editor, two anonymous referees, and conference participants for their comments, Caroline Kouassiaman for outstanding work as project coordinator, and Innovations for Poverty Action for their support on the ground. Financial support from the World Bank's Gender Action Plan and Research Support Budget is gratefully acknowledged.

1. Introduction

Accurate measurement of microenterprise profits and sales is crucial for understanding the dynamics of enterprise growth and the success of policies designed to raise the incomes of the self-employed. However, such data are notoriously noisy. Two indicators of this are the high standard deviation relative to mean seen in cross-sectional profit data, and the relatively low autocorrelation over time seen in panel data. For example, the coefficient of variation in profits among control firms in Banerjee et al. (2009) is 73[1], while McKenzie and Woodruff (2008) report that the percentage change in monthly profits from one calendar quarter to the next in their Mexican data ranges from -97.6 to +4,110. In some cases the data quality is so poor that the authors choose to not use profit or sales data, even though firm performance is a key outcome of interest (See, for example, Dupas and Robinson, 2009). While much of this cross-sectional variation may reflect genuine differences in outcomes among similarly sized firms, and much of this temporal volatility may reflect actual shocks and seasonality experienced by firms, there is a suspicion that much is also due to mismeasurement.

The increasing availability and affordability of computerized data collection through the use of Personalized Digital Assistants (PDAs), Ultra-Mobile Personal Computers (UMPCs), and other such hardware offers one potential new avenue for improving the quality of measurement of microenterprise performance.[2] These devices have been used in many developing countries since the mid-2000s, including in some large-scale surveys such as the Demographic and Health Survey (DHS) in Peru.[3] Some of the advantages and disadvantages of these experiences are discussed in Rosero-Bixby et al. (2005), who note that potential advantages include more accurate and speedier data entry, better handling of skip patterns, the ability to sample on the spot, collecting GPS readings (see also Gibson and McKenzie, 2007), and taking photographs all as part of the same data collection. Caeyers, Chalmers and de Weerdt (2009) show in an experimental setting that computer-assisted personal interviewing results in fewer inconsistencies and less outliers in consumption than paper based questionnaires. Potential disadvantages include costs, the technical expertise needed to develop the electronic questionnaire, difficulties working in non-Roman languages, risk of theft and battery failure, and problems arising from the small PDA screen and interface. Nevertheless, there is a sense that many of these problems will become less severe as technology improves and costs fall over time, making electronic data collection more and more common.

[1] Note Banerjee et al. (2009) calculate profits as revenues less expenses. As shown in de Mel et al. (2009), this results in a much larger standard deviation in profits than directly measuring profits. This is evidenced by Banerjee et al. having a coefficient of variation of 3.8 for sales--still large, but much smaller than for profits.

[2] A second, complementary, avenue is through better questionnaire design. See de Mel et al. (2009) for discussions of how to measure profits.

[3] See for example http://www.macrointernational.com/Information/DataColl/mobile.aspx [accessed July 28, 2009]

In this paper we consider whether the use of consistency checks in PDAs can improve the measurement of microenterprise profits and sales, both in the cross-section and especially in a panel setting. We do this in the context of a quarterly panel of microenterprises in urban Ghana being collected as part of an ongoing randomized experiment. In the cross-section firm owners are queried if they report profits greater than or equal to sales, while in the panel they are queried if the change in profits or sales from one quarter to the next exceeds a predetermined threshold. In practice, the cross-sectional check also served as a second prompt on profits in cases where the firm owner gave sales data but had missing profits data. We are unaware of previous uses of consistency checks in a panel setting, especially with enterprises, and believe that such checks are an innovation that are likely to be adopted by others as electronic data collection becomes more common. We retain both the raw data and any changes after these consistency checks are recorded, allowing us to determine how much difference these consistency checks make.

We find that such checks do reduce the coefficient of variation and increase the autocorrelation of the observations that get revised. The use of the consistency check as a second prompt on profits when the owner had given sales data, but only given a range or no answer for profits, resulted in many of these missing values being corrected. However, only 3-13 percent of the observations in a survey round end up being revised, and the revisions are almost always less than an order of magnitude. In the panel setting, we find only 3-15 percent of the large changes in profits and sales that trigger our consistency checks result in corrections, with firm owners confirming the majority of these large changes as genuine. This shows that much of the high volatility seen in firm sales and profits is driven by shocks and seasonality, not by measurement error. As a result, the consistency checks have very minimal effect on the means, standard deviations, and autocorrelations of sales and profits for the full dataset. As such, it appears that the use of PDAs or other electronic data collection methods for measuring firm profits and sales is unlikely to be justified on the basis of better measurement alone, but if such tools are being used, then we believe including these panel consistency checks is a useful addition.

2. The Survey and use of PDAs

The data are from the Ghana Microenterprise Survey, a quarterly panel survey designed by the authors for the purpose of carrying out a randomized experiment in which small grants will be given to randomly selected enterprises in the cities of Accra and Tema.[4] The broader project is interested in examining differences between male- and female-owned microenterprises depending on whether they work in industries which are dominated by their own gender, or in

[4] These grants were given out following the second and third waves of the survey. However, there appears to be no relationship between the corrections being examined in this paper and receiving a treatment or not. For example, 3.2% of the treated and 2.9% of the untreated (p=0.81) had a profit correction in round 3, and 2.1% of the treated and 2.5% of the untreated had a sales correction (p=0.75). We therefore ignore the treatment in our discussion of the results of consistency checks.

gender-mixed industries. It therefore aimed to work with a representative sample of microenterprises which was stratified by gender and sector.

2.1 Sampling and Survey Design

The sample was constructed as follows. First, enumeration areas (EAs) were selected with probability proportional to the number of households in these EAs according to the 2000 census. We randomly selected 70 EAs in Accra and 30 in Tema. Then, to reduce the costs of listing, we subdivided EAs into equal areas, such that each area would contain approximately 70 to 80 households. This typically required dividing an EA into half or thirds. One of these areas was then randomly selected from each EA. Enumerators went door to door in this area to carry out a screening survey of each household. Households were screened for whether they contained any individual who was aged 20 to 55, and who was self-employed, worked 30 or more hours per week, had no paid employees, and did not own a motor vehicle that was used in the business.[5] The gender and sector of all individuals in a household passing this screen was then recorded. This resulted in screening 7,567 households to identify 3,907 individuals who passed the screen. Only 19.4 percent of these individuals were male, showing the predominance of women among small enterprise operations in urban Ghana. In order to minimize the spillovers from the treatments to be carried out, we did not want to select too many individuals from any given EA who were in the same line of business. We therefore randomly selected up to 5 males in male-dominated and up to 5 males in mixed industries from each EA, and up to 3 females in female-dominated and up to 3 females in mixed industries from each EA to survey, in the process ensuring that only one individual was chosen from any given household.

This resulted in a sample of 907 firms, consisting of 538 females and 369 males. A baseline survey of these firms was conducted in October and November 2008. The firm owners were asked for details of both their firm and their household, with the median firm taking 68 minutes to answer the survey. The key variables collected for the purposes of this paper are firm profits and firm sales. Profits are measured via a direct question, following the recommendations of de Mel et al. (2009). Firm owners were directly asked "after paying all expenses, what were the profits of the business during the month of September 2008? Do not include any income you paid yourself as an expense". Firm owners who replied "don't know" or who refused to answer this question were then asked a follow-up range question, which asked whether their profits were in given ranges. Sales were asked as the total monthly sales from all sources in the past month. Follow-up surveys were then conducted of these same firms in February 2009, May 2009 and August 2009.

[5] These restrictions on employees and not owning a motor vehicle for the business use were made in order to target the grant program being implemented.

2.2 Electronic Data Collection and Consistency Checks

Both the listing exercise and the survey itself were carried out using PDAs. Specifically, the hardware used was the HP iPAQ Business Navigator (614c), which included a digital camera and GPS capability. At the time of the survey, retail price was US$350 per unit. The survey was programmed using PocketSurvey, which was used by Oxford University under an annual multi-unit license. At the time of writing the cost of this software is £5000 + VAT for a single user, £250+VAT per additional handheld unit, and £1000+VAT for each additional desktop license. Running a standalone survey with two desktop licenses and twenty handheld units therefore would cost £12,300 (approx. US$20,000).[6] Programming the baseline survey took approximately 75 hours of researcher time, while programming the follow-up surveys which retained many of the same questions each round took approximately 20 hours per round.

We programmed consistency checks for sales and profits in both the cross-section and in the panel. In choosing when to trigger a consistency check we had to trade off the potential for improved data against the cost of lengthening the questionnaire time for firms. We therefore chose to keep the number of checks relatively parsimonious, restricting them to the key panel variables and restricting the number of cross-checks. In the cross-section, we carried out two checks on profits and sales. The first was triggered if reported profits were greater than reported sales, while the second was triggered if reported profits equaled reported sales and reported expenses exceeded zero. If the check was triggered, firm owners were told, for example, "You told us sales were 100 and profits were 120. Is it really true that your profits exceeded your sales last month?" They were then asked to confirm what their profits and sales actually were.

One could consider comparing directly reported profits to directly reported revenues less expenses. However, there are a number of reasons why the two need not be equal over a one month period (see de Mel et al. 2009), chief among them being a mismatch between when expenses are made and when the revenues associated with those expenses are realized. Our less strict consistency check is not subject to this problem, and should capture mistakes. The checks were intended to catch keypunch and misreporting errors, as well as cases in which the concepts of sales and profits have been mixed up or treated as the same thing by the owner. In practice however, PocketSurvey treats missing values as infinity (similar to Stata), so that if profits are missing and sales are not, then this check will also be triggered.[7]

In the panel our consistency check aims to determine whether large changes from one quarter to another in monthly sales and profits are genuine or mismeasurement. We therefore programmed PocketSurvey to pull in the previous round's sales and profits data, and calculate the absolute and percentage change in monthly profit and in monthly sales. A consistency check

[6] Pricing obtained from PocketSurvey on August 4, 2009. There is also a 15% annual maintenance and support charge for subsequent years.

[7] It would also be triggered if both profits and sales were missing, but expenses were reported as zero.

was then triggered if the percentage change was less than -33.3 percent or greater than +50 percent in the quarter and if the absolute change in profits was at least 20 cedi and absolute change in sales at least 50 cedi.[8] The absolute criteria prevented firms with very low profits or sales from having to answer consistency check questions when the actual change was very small. The percentage change thresholds were chosen for symmetry of absolute change, and were based on approximately the 25th and 75th percentiles of changes in profits seen in the microenterprise panel data from Sri Lanka used in de Mel et al. (2009). They also turned out to be close to these same percentiles for changes in profits in the raw (uncorrected) data in Ghana.

Firms which were triggered in the panel check were then told, for example, "in the last survey you told us your January sales were 100, and you have just told us your April sales were 800. Have your sales really increased this much?" They were then asked to confirm both the previous round's and current round's sales.[9] One might question whether there is more measurement error in asking firms to recall sales or profits from four months ago compared to one month ago. De Mel et al. (2009) compare sales reported over four month and one month intervals in Sri Lanka and find a Spearman correlation of 0.65, with mean sales 10 percent lower when asked with the longer recall. Since we are only asking for this recall when the percentage change in sales or profits is much larger, we believe that any measurement error introduced by recall is second order compared to the potential measurement error triggering the consistency check. That is, a firm owner might not recall correctly whether January sales were 100 versus 80, but should remember whether they were one of these versus 800.

These checks were intended to be implemented from round 2 onwards. However, a programming error with skip patterns meant that the check on changes in profits was only implemented starting in the third round. We therefore have the cross-sectional consistency checks in all rounds, the panel checks occurring in round 2 onwards for sales, and the panel checks occurring in round 3 onwards for profits.

Note that it is the panel consistency checks which rely most on the PDA for easy implementation. It order to do this with a paper-based survey, one would have to have a paper page for each firm which listed its previous profits and sales, and then rely on the enumerator to calculate the absolute and percentage changes. This is certainly feasible when only one or two key questions are being checked in the panel, but still likely to be prone to more errors, and would become less feasible if one wished to employ panel consistency checks for a wider range of variables. In contrast, the cross-sectional consistency checks wouldn't be very difficult to implement in a paper-based survey, although the PDA still simplifies this. In the case of missing

[8] One Ghana cedi was approximately equal to one US dollar at the time of the baseline survey.

[9] Note that enumerators did not know the previous round's figures unless the check was triggered. While one might be concerned that enumerators might simply answer that any errors were genuine in order to speed up their interview work, our enumerators were of high quality with professional survey experience, and supervisor checks and direct field observation did not reveal any instances of this occurring.

values for profits, the cross-sectional consistency check essentially served as a second prompt on profits, but with reference to sales being reported. While enumerators were told to push to obtain an exact figure for profits, in the first rounds of the survey it seems they sometimes relied on a range response only, which triggered a missing value in the check. The consistency check then essentially caused them to press harder for an exact figure on profits. In this case the PDA effectively acts as a constraint on enumerators not pressing hard enough for a figure. It is plausible that more intensive enumerator training could rectify this, but the PDA helps automate and simplify the process of re-prompting for missing values in a case where the firm owner is willing to provide related information.

One potential concern with using these consistency checks is how the firm owners will react when queried on their numbers. Enumerators were instructed to ask these clarifications in a non-confrontational manner, and to make clear that a reason for an error could be that the data were incorrectly recorded. In practice we did not experience problems in the field with firm owners reacting negatively to these consistency checks. The enumerators were all professional enumerators, many with masters degrees and significant experience in surveying. Coupled with oversight and field checks, we believe that the enumerators did implement the questionnaires and checks in the manner intended.

3. Results

3.1 How frequent are revisions in profits and sales because of consistency checks?

Table 1 summarizes the frequency of consistency checks performed each round, and the number of changes made by the firm owner in response to these consistency checks. We report both the total number of observations changed, and the number changed from missing to a non-missing value. The final column then calculates the percentage of changes where the consistency check results in an increase in the reported value of the variable in question to give a sense whether changes are all occurring in one direction or not.

Consider first the cross-sectional checks. The number of firms triggering this check was large in the baseline, primarily as a result of a high incidence of firms with missing values for the profits question (many of whom answered a range question instead). When these firms were then queried because the handheld viewed a missing value for profits as exceeding sales, 66 of these firms then did give a profits number. This also occurs with high frequency in the second round of the panel, but by the time of the third round most firm owners were reporting profits in the first question, before this check. In addition, slightly fewer firm owners report profits equal to sales in the later rounds of the survey, suggesting some learning of concepts is occurring.[10] As a result,

[10] This idea that it can take households (and researchers) a few rounds of the survey to understand the concepts and how to measure them is seen in the monthly Thai data collected by Samphantharak and Townsend (2009), who drop the first four months of data in their analysis to allow time for the logistics and methods to stabilize. This is not possible in most studies, so the question is how consistency checks can be used to help correct misunderstandings.

only 2-3 percent of firms have changes as a result of this consistency check in rounds 3 and 4, compared to 10-12 percent in rounds 1 and 2.

For profits then, the majority of the changes from the cross-sectional consistency check come from changing a missing value to a non-missing value. This is not the case for sales, where a missing value would only trigger a check if both profits and sales were missing (and thus equal). The majority of changes in sales instead came from firm owners who had missing profits now reporting profits, and in the process of reporting profits, also revising their sales figure. There are a small number of firms who report sales and profits to be the same, possibly as a result of not understanding the concept or of the translation being similar for both terms. The consistency check picks up these cases too, and results in changes. As one would expect, there is a tendency for the cross-sectional checks to lead to a downward revision of profits and upwards revision of sales, although this pattern is stronger in some rounds than others.

More firms are subjected to the panel checks than the cross-sectional checks, with just over one-third of firms given the profits check, and almost half of the firms receiving the sales check in a given round. In the round 2 data for sales firms are subject to two panel checks – a check of round 2 data compared to round 1, and a check of round 2 data compared to round 3. This double-panel check also occurs for rounds 3 for both profits and sales, while round 4 data currently have only a single panel check of the round 3 to round 4 change.

The most notable result with the panel checks is how few firms change their responses when prompted for the check. For example, 287 firms trigger a consistency check due to the size of their profits change between round 2 and round 3. This results in changes for 10 of the round 2 profits figures and 9 of the round 3. There are more changes for round 1 sales data on the basis of the round 2 checks, but this still only accounts for 6.3 percent of the overall sample. In all other rounds the panel checks are resulting in fewer than 2 percent of observations being changed. This is noteworthy because firm owners are thus confirming that the amount of volatility seen in their sales and profits data is largely genuine, and not the result of measurement error.

3.2 Are larger percentage changes more likely to be revised in the panel checks?

One potential concern with the panel consistency checks is that the mere act of questioning whether the change is genuine may cause some respondents to revise their profits and sales figures, even when no error was made. The small proportion of changes made in response to these consistency checks suggests that this is unlikely to be a widespread phenomenon. We then would like to know whether it is the larger changes that are more likely to be revised or not. Since Table 1 showed that there were only a small number of revisions made as a result of the panel checks, we choose to examine this question in a fairly simple way, through t-tests.

Looking at the distribution of the changes in profits between rounds 2 and 3 for observations triggering the profit check, we see that the upper and lower quartiles are approximately +150% and -60%. We therefore classify as "relatively large" percentage changes which are greater than or equal to +150% or less than or equal to -60%. Table 2 then reports the proportion of relatively large versus relatively small percentage changes which led to a revision in the panel check, along with p-values from t-tests of equality. For profits we see 4.6% of the larger changes lead to a revision in the round 3 panel check, compared to 2.7% of the smaller changes, although the difference is not statistically significant. For the sales the differences are statistically significant, with 5-15 percent of the relatively large changes resulting in a revision, compared to 0-5 percent of the relatively smaller changes. Thus while relatively larger changes are more likely to result in a revision, firm owners are confirming that the vast majority of the large changes in profits and sales measured are indeed genuine. This highlights the extreme volatility (due to seasonality, shocks, and other factors) of enterprise incomes.

3.3 What sort of revisions to profits and sales do firm owners make, and who is more likely to revise?

Next we examine what types of revisions are made when firm owners revise sales and profits. One obvious source of errors is cases where the enumerator mis-entered the data, either by entering or omitting zeros. However, such corrections are not the principal reason for the revisions being made – only 7.3 percent of the profit revisions and 2.8 percent of the sales revisions over the four rounds involved simply adding or removing zeros. More generally, few of the revisions made involve an order of magnitude correction to the raw data. Only 12 percent of the profit revisions and 11 percent of the sales revisions result in a correction made to the raw data which is a factor of ten or larger (or of 0.1 or smaller). Taking the maximum of the ratio of the corrected data to the raw data and its inverse, the median of this ratio is typically 2 to 3. Thus a more typical correction involves the firm owner revising a reported value of 60 to 30, or from 200 to 600. Some of these changes may represent single-digit entry errors on the part of the enumerator, while others may involve errors in reporting by the firm owner. The corrections are still relatively sizeable, but because they are not an order of magnitude, they would be difficult for standard ex-post error checking to detect.

Next we explore whether there are certain individual characteristics which are associated with individuals being more likely to correct errors. We can then think of several factors which may make individuals more likely to make mistakes, which can then guide our choice of variables. Profits and sales may be mis-reported intentionally, perhaps due to fear of taxes or theft, in which case we might expect firms with a tax identification number to be more likely to make internal consistency errors as a result. People might also make mistakes because they do not understand accounting, are not good with numbers, or because they are disorganized in their record-keeping. We include controls for the education of the owner and whether or not the firm keeps accounts for this purpose. We also control for gender of the owner, since if women

typically control the household finances, they may have more difficulty separating household and business expenses, leading to errors in recalling profits. Mistakes may also arise because firms operate in a sector where it is harder to keep track of sales and expenses because of high turn-around. We therefore control for whether or not the firm is in trade and retail, which is the sector with most rapid turnover. Finally, we control for age for the firm, since owners of relatively new firms may have had less time to develop a sense of what profits and sales are.

Given the low frequency of corrections we pool the different rounds to create a dependent variable of ever making a correction over any round, and examine its association with individual characteristics in Table 3. We find that owner and firm characteristics are not strong predictors of which firm owners will make corrections, with none of the coefficients statistically significant or large in magnitude. Indeed, we cannot reject the null hypothesis that these firm and owner characteristics are jointly unrelated to whether a firm owner makes a correction. This continues to hold true after controlling for baseline interviewer fixed effects. One reason for the significance of interviewer effects is that it appears that some interviewers were more likely to not push for an exact answer to profits in the baseline and instead record a range, which the error check then corrected. However, even when we consider error checks other than for those correcting missing values, interviewer effects are still jointly significant. It is unclear whether this reflects some interviewers making more errors than others, or some interviewers pushing the firm owner more to understand inconsistencies in the cross-section and panel checks when they arise.

We can also examine whether it is the same people making corrections in the panel each round, or whether it is different people each time. We find only 4 percent of firm owners making a panel correction in sales make it because of both the round 2 and round 3 consistency checks, and the people making the sales revisions because of the round 4 panel check were all different from those who made it as a result of the round 3 check. Only one out of the 27 individuals making a profits revision because of either the round 3 or the round 4 profits panel checks made a revision because of both checks. The vast majority therefore are only making it in only one of the panel checks, not both. Thus it is not just a few individuals who systematically make and correct errors, but rather appears to be error corrections are randomly distributed across firm owners.

3.4. How does making these changes affect the distributions of sales and profits?

Finally we examine what the impact of these consistency checks is on the distribution of profits and sales. Panel A of Table 4 begins by summarizing how the corrections made affect the mean, standard deviation, coefficient of variation, and autocorrelation of the observations for which corrections were made. Recall that one of the main effects of the consistency checks was to revise missing values to non-missing. However, to compare the distributions of the same observations, we only consider the observations which were non-missing in the raw data here.

We see the corrections result in quite large changes in the data which is corrected, with the corrected data usually having a lower coefficient of variation and higher autocorrelation. For example, the corrections increase the correlation of the corrected round 2 and round 3 sales from 0.319 to 0.791, and lower the coefficient of variation of the round 3 sales data from 4.0 to 1.8. These changes are consistent with the corrections reducing the noise in the data. Moreover, comparing the autocorrelations in Panel A to those in Panel B, which show the full sample, we see the autocorrelation of the corrected data becoming closer to that of the overall sample. We view this as suggesting the corrections are indeed taking the corrected observations closer to the distribution of the underlying data.

However, because only a small number of observations are changed each period and the changes are not orders of magnitude changes, we see in Panel B that the consistency checks have limited effect on the distribution of the overall sample. The coefficients of variation are generally equal for the raw and corrected data. There are small improvements in the autocorrelation, but only between 0.01 and 0.05, with larger improvements for sales than profits. Such improvements are useful in improving the power of experiments using this data to detect a given effect size, but do not suggest overwhelming improvements from the use of electronic consistency checks. Consider, for example, the sample size needed to detect a 20 percent increase in either profits or sales because of some intervention, using five rounds of a panel. Holding the coefficient of variation fixed, the improvement in autocorrelation seen in the round 2 sales data would reduce the sample size needed to detect such a treatment effect from 958 to 892, a reduction in sample size of 66.[11] In contrast, the 0.009 modal increase in the autocorrelation seen for profits would reduce the sample size needed to detect a 20 percent treatment effect by only 8 firms.[12] The larger increase in effective sample size in our context comes from the checks replacing missing values with non-missing values – a total of 124 observations or 31 on average per round were gained in this way.

Given survey costs of approximately $20 per firm per round, the savings from these gains in sample size would be approximately $3,900-6,600 in survey costs. Note these savings would greatly exceed the marginal costs of adding consistency checks to a survey that is taking place with PDAs. When comparing PDAs to paper-based surveys, the larger savings in costs come from not having to photocopy questionnaires each round, and from not having to pay for data entry. The use of handhelds and software may therefore still pay for itself, especially when amortized over multiple rounds or several projects, but the savings from greater power to detect given effect sizes due to increased accuracy are likely to be only a small part of such calculations. Trimming outliers is an alternative, cheaper way of increasing the autocorrelation

[11] This is calculated using the sampsi command in Stata, assuming ANCOVA as the method, that the treatment occurs after the second round of the data and is given to half the sample, with a baseline mean of 500 and standard deviation of 825.

[12] Based on a baseline mean of 140, standard deviation of 180, and varying the autocorrelation from 0.563 to 0.572 as in round 3 of the profits corrections.

and decreasing the coefficient of variation. However, the number of large changes which appear to represent actual changes suggests that trimming may be eliminating real variance rather than noise. For example, one approach to dealing with extreme variation is to trim the top 5 percent of percentage changes in profits: however, in our data only 4 out of 72 observations in the top 5 percent of raw profit percentage changes are confirmed by firm owners as genuine errors and corrected during our checks.

4. Conclusions and Discussion

We implemented cross-sectional and panel consistency checks with the aim of improving measurement in microenterprise profits and sales. There were two main effects of these checks. The first was a reduction in the number of missing values for profits, as firms who were queried again about profits after telling us sales often gave an explicit number for profits where previously they had not answered, or answered only with a profit range. These appears to be largely an effect of pushing particular enumerators to push harder for exact figures on profits, rather than of the questions inducing particular types of firms to answer who wouldn't before. Second, the checks did reduce the coefficient of variation and increase the autocorrelation for the observations which were corrected. However, the overall effects were fairly modest, since few observations were changed and the changes were not order of magnitude changes.

One of the key challenges for attempts to improve measurement in surveys is knowing whether changes in measurement are actually improvements or not. This is particularly the case when measuring the profits and sales of microenterprises, for which no gold standard administrative data is available for comparison. Despite this, we judge the consistency checks to have improved measurement in our survey. The argument for improvement is clear when missing values now become answered, with answers that appear reasonable in light of the reports of other similar businesses. For the revisions to already answered data, we consider the fact that the revisions pick up obvious inconsistencies and that firms realize, and correct, these inconsistencies, to validate the cross-sectional checks. The panel checks improve the autocorrelation of the data and bring the distribution of the corrected data closer to that of observations where corrections were not made. In the absence of an alternative benchmark for judging success, we consider this as evidence of an improvement in measurement.

The final issue is then whether the improvements generated by these consistency checks are in of themselves enough reason for researchers to switch from paper surveys to electronic data capture when surveying microenterprises. Based on the evidence here, we conclude that they are not. The potential improvement in power from reducing large changes in profits and sales is more limited than we a priori expected, since firm owners validate the majority of large changes as genuine. This provides further evidence for the volatility of income experienced by the poor when running businesses, a point also noted from recent studies of financial diaries of the poor (Collins et al. 2009). This is not to say that a shift to electronic data collection might not

be justified on other grounds – in some cases the cost may be less than the costs of photocopying and data entry for example – just that the ability to carry out consistency checks of the nature studied here is unlikely to be the prime reason for the shift.

References

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster and Cynthia Kinnan (2009) "The miracle of microfinance? Evidence from a randomized evaluation", Mimeo. MIT.

Caeyers, Bet, Neil Chalmers and Joachim de Weert (2009) "A Comparison of CAPI and PAPI through a Randomized Field Experiment", Mimeo. University of Oxford.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford and Orlanda Ruthven (2009) *Portfolios of the Poor: How the World's Poor Live on $2 a day.* Princeton University Press, Princeton, NJ.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2009) "Measuring microenterprise profits: Must we ask how the sausage is made?", *Journal of Development Economics* 88(1): 19-31

Dupas, Pascaline and Jonathan Robinson (2009) "Savings constraints and microenterprise development: Evidence from a field experiment in Kenya", Mimeo. UCLA.

Gibson, John and David McKenzie (2007) "Using the Global Positioning System (GPS) in household surveys for better economics and better policy", *World Bank Research Observer*, 22(2): 217-41.

McKenzie, David and Christopher Woodruff (2008) "Experimental evidence on returns to capital and access to finance in Mexico", *World Bank Economic Review* 22(3): 457-82

Rosero-Bixby, Luis, Jeisson Hidalgo-Céspedes, Daniel Antich-Montero and Mitchell A. Seligson (2005) "Improving the quality and lowering costs of household survey data using Personal Digital Assistants (PDAs). An application for Costa Rica", Paper presented at the 2005 Population Association of America Annual Meetings, March 31-April 2, Philadelphia, PA.

Samphantharak, Krislert and Robert Townsend (2009) "Households as Corporate Firms: Constructing Financial Statements from Integrated Household Surveys", *Econometric Society Monographs* (No. 46)

.

Table 1: Frequency of Consistency Checks and Revisions by Survey Round

Variable	Survey Round	Number of Observations	Check	# Obs. Triggered	# Obs. Changed	# Obs. Changed from Missing	% of Changes not from missing	% of all Obs. which get changed	% of changes which are positive
Profits	1	804	Cross-section	220/147	93	66	29.0	11.6	14.8
Profits	2	788	Cross-section	94/76	73	45	38.4	9.3	64.3
Profits	2	788	Round 3 panel	287	10	0	100.0	1.3	50.0
Profits	2	788	All checks	367/349	82	45	45.1	10.4	62.2
Profits	3	749	Cross-section	59/34	15	7	53.3	2.0	0
Profits	3	749	Round 3 panel	287	9	0	100.0	1.2	77.8
Profits	3	749	Round 4 panel	314	6	0	100.0	0.8	50.0
Profits	3	749	All checks	504/479	29	7	75.9	3.9	40.9
Profits	4	730	Cross-section	48/36	20	6	70.0	2.7	7.1
Profits	4	730	Round 4 panel	314	12	2	83.3	1.6	70.0
Profits	4	730	All checks	349/337	29	6	79.3	4.0	34.8
Sales	1	867	Cross-section	220	68	3	95.6	7.8	55.4
Sales	1	867	Round 2 panel	448	55	0	100.0	6.3	83.6
Sales	1	867	All checks	561	112	3	97.3	12.9	71.6
Sales	2	805	Cross-section	94	78	2	97.4	9.7	55.3
Sales	2	805	Round 2 panel	448	11	0	100.0	1.4	63.6
Sales	2	805	Round 3 panel	355	9	0	100.0	1.1	66.7
Sales	2	805	All checks	604	93	2	97.8	11.6	52.7
Sales	3	783	Cross-section	59	13	0	100.0	1.7	92.3
Sales	3	783	Round 3 panel	358	9	0	100.0	1.1	66.7
Sales	3	783	Round 4 panel	297	3	0	100.0	0.4	33.3
Sales	3	783	All checks	520	23	0	100.0	2.9	73.9
Sales	4	741	Cross-section	48	23	0	100.0	3.1	69.6
Sales	4	741	Round 4 panel	297	6	0	100.0	0.8	83.3
Sales	4	741	All checks	318	28	0	100.0	3.8	71.4

Notes:

Number of observations is number of non-missing values after corrections have been made in response to consistency check

In cases where two numbers are given for number of observations triggered, the first indicates the total number triggered, while the second is the total number triggered among the final sample of non-missing values for this variable.

Percentage of changes which are positive excludes changes from a missing value to a non-missing value.

Table 2: Are more extreme changes in profits or sales more likely to be revised?

Panel consistency check	# Obs. Triggered	Proportion of observations revised When change is relatively small	When change is relatively large	T-test p-value
Round 3 profits check	287	0.022	0.046	0.273
Round 4 profits check	314	0.025	0.033	0.650
Round 2 sales check	448	0.053	0.149	0.005
Round 3 sales check	355	0.000	0.085	0.0002
Round 4 sales check	297	0.007	0.045	0.047

Note: relatively small changes are for percentage changes of -33.3 to -59.9%, or +50 to 149.9%
Relatively large changes are percentage changes of -60% or less, or +150% or greater.

Table 3: Are certain firm owners more likely to make corrections?

Marginal effects from probit estimation

	Ever correct Profits?		Correct Profits from Missing		Ever correct Sales?	
	(1)	(2)	(3)	(4)	(5)	(6)
Female owner	0.0435	0.0468	0.0373	0.0328	0.0396	0.0226
	(1.378)	(1.483)	(1.478)	(1.130)	(1.200)	(0.658)
Years of education	-0.00547	-0.00296	-0.00382	-0.00398	-0.00368	-0.00249
	(-1.298)	(-0.681)	(-1.203)	(-1.034)	(-0.806)	(-0.543)
Firm is in trade/retail	-0.0388	-0.0278	-0.0348	-0.0319	-0.0273	-0.0315
	(-1.281)	(-0.926)	(-1.461)	(-1.193)	(-0.858)	(-0.981)
Firm owner keeps accounts	0.0151	-0.0103	0.0386	0.0376	0.0438	-0.0231
	(0.484)	(-0.305)	(1.520)	(1.243)	(1.345)	(-0.632)
Firm has a tax id	0.0191	-0.00316	0.0271	0.00319	0.0654	0.0663
	(0.459)	(-0.0705)	(0.800)	(0.0830)	(1.488)	(1.290)
Age of firm	-0.00140	-0.000122	-0.00287	-0.00244	-0.000158	-0.000142
	(-0.685)	(-0.0591)	(-1.619)	(-1.161)	(-0.0732)	(-0.0676)
Interviewer fixed effects	no	yes	no	yes	no	yes
Chi-squared p-value						
Firm and owner characteristics	0.337	0.584	0.087	0.395	0.425	0.708
Interviewer effects	n.a.	0.000	n.a.	0.000	n.a.	0.000
Observations	847	834	847	684	847	838

Robust z-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Chi-squared tests null hypothesis that coefficients on firm & owner characteristics jointly zero.

Table 4: How much difference do the consistency checks make?

Panel A: Summary Statistics for Data Where a Correction Was Made and Raw Data was Not Missing

		Raw Data				Corrected Data			
	# Obs.	Mean	S.D.	C.V.	Pearson/Spearman Autocorrelation	Mean	S.D.	C.V.	Pearson/Spearman Autocorrelation
Round 1 profits	27	244	381	1.6		78	171	2.2	
Round 2 profits	37	104	251	2.4	0.082/0.356	108	150	1.4	0.273/0.184
Round 3 profits	22	129	162	1.3	0.054/0.308	83	85	1.0	0.054/0.315
Round 4 profits	23	102	80	0.8	0.345/0.435	113	130	1.2	0.694/0.562
Round 1 sales	109	381	1241	3.3		474	995	2.1	
Round 2 sales	91	622	2256	3.6	0.110/0.180	328	583	1.8	0.432/0.549
Round 3 sales	23	3021	12074	4.0	0.319/0.403	944	1700	1.8	0.791/0.510
Round 4 sales	28	309	454	1.5	0.864/0.246	488	951	1.9	0.745/0.457

Panel B: Summary Statistics for the Full Data were Raw Data Was Not Missing

		Raw Data				Corrected Data			
	# Obs.	Mean	S.D.	C.V.	Pearson/Spearman Autocorrelation	Mean	S.D.	C.V.	Pearson/Spearman Autocorrelation
Round 1 profits	737	132	244	1.8		126	234	1.9	
Round 2 profits	742	144	317	2.2	0.235/0.450	144	314	2.2	0.258/ 0.441
Round 3 profits	742	136	185	1.4	0.563/ 0.668	135	184	1.4	0.572/ 0.671
Round 4 profits	724	142	181	1.3	0.677/0.649	142	182	1.3	0.686/0.653
Round 1 sales	858	400	758	1.9		423	793	1.9	
Round 2 sales	799	514	859	1.7	0.497/ 0.474	505	825	1.6	0.546/ 0.560
Round 3 sales	766	561	841	1.5	0.667/0.641	567	834	1.5	0.708/ 0.666
Round 4 sales	739	605	915	1.5	0.716/0.698	612	927	1.5	0.737/0.712

Note: Full data trims observations above 10,000 cedi which are likely currency errors.