# Remittances and Rashomon

## Randall Akee and Devesh Kapur

### **Abstract**

Utilizing a novel data set on remittance data for India that matches household surveys to administrative bank data, we investigate the differences in self-reported and actual deposits to Non-Resident Indian (NRI) accounts. There is a striking difference between the perceived and actual frequency, as well as the amount of deposits, to NRI accounts. Our results indicate the presence of non-classical measurement error in the reporting of remittances in the form of deposits to NRI accounts. As a consequence, regression analyses using remittances as an explanatory variable may contain large upward biases instead of the usual attenuation of results under classical measurement error. Instrumental variables estimates are no better; the estimated coefficients from these regressions are more than three times the size of the OLS regression results. The results point to the need to more carefully check the accuracy of the international remittance flows. The wide discrepancies in the Indian case could be both because of inaccuracies in the household survey as well as mis-classification of the Balance of Payment data with some fraction of reported remittances being disguised capital flows (and hence likely to be less stable) rather than current account flows for family maintenance.

Keywords: Remittances; Measurement Error; Migration.



Working Paper 285 January 2012

#### Remittances and Rashomon

Randall Akee
Tufts University Economics Department
randall.akee@tufts.edu

Devesh Kapur

Director, Center for the Advanced Study of India

University of Pennsylvania

January 2012

Acknowledgements: This project was conducted in collaboration with the Reserve Bank of India. We would like to thank the institution and in particular A. P. Gaur, Sharad Kumar, Vivek Kumar, Dr. Balwant Singh and Dr. Charan Singh for their singular commitment and assistance in making available the data for this research. The authors would also like to thank Jenny Aker, Catia Batista, Lakshmi Iyer, Arvind Subramanian and participants at the Harvard University Kennedy School of Government Conference on Migration and Development, seminar participants at the Reserve Bank of India and the Indian Statistical Institute, UC Berkeley Labor Seminar and the Center for Global Development for their helpful comments. This research was finalized while Randall Akee was a visiting fellow at the Center for Global Development and is grateful for the center's generosity in terms of resources and valuable input. Any errors, omissions or oversights are ours alone.

CGD is grateful for contributions from the John D. and Catherine T. MacArthur Foundation in support of this work.

Randall Akee and Devesh Kaput. 2012. "Remittances and Rashomon." CGD Working Paper 285. Washington, D.C.: Center for Global Development. http://www.cgdev.org/content/publications/detail/1425889

Center for Global Development 1800 Massachusetts Ave., NW Washington, DC 20036

> 202.416.4000 (f) 202.416.4050

The Center for Global Development is an independent, nonprofit policy research organization dedicated to reducing global poverty and inequality and to making globalization work for the poor. Use and dissemination of this Working Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

www.cgdev.org

The views expressed in CGD Working Papers are those of the authors and should not be attributed to the board of directors or funders of the Center for Global Development.

#### I. Introduction

Remittances to India have grown dramatically in the past two decades. According to official Reserve Bank of India (RBI) data remittances have grown enormously from \$2.1 billion in 1990-91 to \$53.9 billion in 2009-10.1 The share of private transfer receipts in India's GDP rose from 0.7 per cent in 1990-91 to 3.6 per cent in 2009-10. During this period there has been a concerted shift in the sending of remittances from informal to more formal channels. Overseas remittance inflows to India come primarily through two channels:

- i) Inward remittance towards family maintenance.
- ii) Non-Resident Indian (NRI) deposits schemes with the banks in India.<sup>2</sup>

The first channel was previously dominated by informal methods such as *hawala*, but Western Union and other forms of wire transfers are the most common means today - accounting for almost half of the remittance flows to India. There has been a similar growth in the flow of funds through the NRI accounts as well. These NRI accounts allow the immigrant abroad to deposit funds to a bank account held in India that can be accessed by specifically designated parties (typically household members). Local withdrawals from NRI bank deposits previously accounted for around 30 per cent of remittance flows in the late 1990s; in the last decade, however, that amount has grown to between 45-50 percent. According to estimates by the RBI, in 2009-10 total private transfer inflows into India were \$53.9 billion. Of this \$29.4 billion was "inward remittances for family maintenance" and \$23.6 billion were "local withdrawls from NRI deposits."

Other developing countries are increasingly turning to these overseas bank accounts as a more secure and efficient method of securing remittances; for instance, the Philippine Central Bank has made a concerted effort to encourage the use of Philippines Payment and Settlement System (PhilPass) for remittances from overseas Filipinos.

As the increase in formality of remittances flows has increased in India, we would expect the accuracy of reporting both on an aggregate and individual level to improve. Unfortunately, at the aggregate level at least, there appears to be a large discrepancy between the official RBI

<sup>&</sup>lt;sup>1</sup> Official Indian data is reported by Financial Year which is from April-March. Thus the data for 2009-10 is from April 1, 2009 to March 31, 2010.

<sup>&</sup>lt;sup>2</sup> While inflows from overseas Indians for deposits in the NRI deposit schemes are treated as capital account transactions, funds domestically withdrawn from the Non-Resident (External) Rupee Account [NR(E)RA] and Non-Resident Ordinary (NRO) Rupee Account are included in the current account as private transfers. NRI deposits in India grew from \$17.2 billion in March 1995 to \$47.9 billion in March 2010.

<sup>&</sup>lt;sup>3</sup> RBI Bulletin, May 2011, Table 9.

Balance of Payments data on remittances to India and aggregated amounts on nationally representative household surveys conducted by the National Sample Survey (NSS).

In principle the aggregation of the micro-household data at the all-India level should be equal to the macro-BOP data – or at least close to it. In reality there is huge difference – by an order of magnitude.<sup>4</sup> The NSS 64th Round conducted between July 2007 – June 2008 focused on "Migration in India" and total remittances from those whose present place of residence was "another country" was Rs 167.06 billion.<sup>5</sup> The macro-numbers from the RBI put the figure for remittances for 2007-08 at Rs 1640.17 billion.<sup>6</sup> While the RBI's data is from April 2007-March 2008 and the NSS covers July 2007-June 2008, the fact that remittances have been growing steadily should have led the aggregate NSS data to be greater (not less) than the RBI data. There is a possibility that the NSS question on remittances "whether sent remittances during the last 365 days" could imply that the aggregation should be done over 2006-07 instead of 2007-08; but even then the RBI data for remittances in 2006-07 is Rs. 1331.12 billion – eight times the NSS aggregate. <sup>7</sup> Researchers have found consistent under reporting in household surveys relative to balance of payment figures in other countries as well, but none have been as large as in the case of India (Andriasik, 2005; Acosta et al, 2006; Shonkwiler et al, 2011).

Accurately measuring remittances flows is particularly important in developing countries where the flows may account for a large proportion of household income. In India, the BOP estimates would indicate that remittances are approximately 4% of GDP, whereas the NSS household aggregate measure would put it only at 0.4% of GDP. The large margin of error is a big area of concern for policymakers. Estimates of income mobility, inequality and growth will all be affected by this mis-measurement. The results point to the need to more carefully check the accuracy of the international remittance flows. The wide discrepancies in the

<sup>&</sup>lt;sup>4</sup> In India the difference between NSS aggregates and macro numbers are not unique to remittances. There has been growing concern, for instance at the widening discrepancies between NSS consumption figures and national income accounts. In 1972-73 consumption measured by the NSSO surveys amounted to 87% of consumption estimated by national accounts (which measures GDP). This proportion of consumption captured by the NSSO has steadily declined over the years, to 48.8% in 2005-05 and to just 43% as per the most recent 2009-10 survey. But even then the ratio is double, not an order of magnitude.

<sup>&</sup>lt;sup>5</sup> Appendix Table 14, p. A-57, NSS Report No. 533 (64/10.2/2).

<sup>&</sup>lt;sup>6</sup> RBI Bulletin March 2010, Statement 5, p. 588.

<sup>&</sup>lt;sup>7</sup> A different way to cross-check data would be to compare the numbers for remittances *outflows* from a country to India provided by that country with the numbers for remittance *inflows* into India from that country provided by Indian authorities. Earlier RBI estimates of the geographical break-up of remittance inflows put the share of North America between 30 – 35 per cent in one report and 44 per cent in another (RBI, 2006a, RBI 2006b). The Appendix Table 1 compares the estimates of remittance flows from the United States to India from a recent report from the Congressional Budget Office with corresponding figures from the RBI for remittance inflows from North America. The discrepancy is almost a factor of five. If we grant that the RBI data is for North America and not just for the US, the difference would still be at least a factor of four since the Indianborn population in the US is at least four times that in Canada and an important factor underlying the large increase in remittances to India in the last decade appears to have been the influx of Indian IT-workers to the US through the H1-B visa program.

Indian case could be both because of inaccuracies in the household survey as well as misclassification of the Balance of Payment data with some fraction of reported remittances being disguised capital flows (and hence likely to be less stable) rather than current account flows for family maintenance. At the microeconomic level, mis-reporting at the household level may significantly affect attempts to estimate the effect of remittances on various types of household investment and consumption expenditures.<sup>8</sup>

While the data suggests that there is a relatively large discrepancy between the macro and micro aggregates for total remittances in India, we focus in this study on a single verifiable channel of remittance flow to India via the NRI accounts. The withdrawals from deposits by immigrants abroad to NRI accounts comprise almost half of all remittance flows to India. Restricting our analysis to the NRI accounts allows us to use administrative data (RBI data on NRI credits and debits) to verify self-reported amounts on household surveys. In this study, a random sample of Non-Resident Indian (NRI) households in four Indian States (Kerala, Gujarat, Maharashtra, Punjab) are surveyed. Households are asked to report in a survey the amount of remittances received annually by several different remittance channels. This household survey is then linked to administrative data from the Reserve Bank of India for these same RBI accounts. To our knowledge, this is the first research which has a matched data set between administrative and survey data on remittances. Because of time and cost, the administrative data is provided for only a subset of the original household survey respondents. However the matched data is representative and provides several important insights into the nature of measurement error of this remittance channel.

We find that the differences between the self-reported frequency of remittances and actual deposits are large; the modal value for self-reports is 12 deposits per year while the administrative data indicate just 2-3 deposits per year. Second, the average amount remitted in a year as self-reported is 345,938 rupees, while the actual average amount deposited per year is 677,269 rupees. There is strong evidence that the mis-reporting displays mean-reversion; individuals with above average levels of remittances tend to underreport the amount that they receive and those with below average remittances tend to over report the amount that they receive. This non-classical measurement error entails complications for regression analysis.

We also examine the role that household characteristics, specifically human capital variables, play in reporting error in remittances. It appears that more highly educated households tend to be more accurate in their survey responses. Additionally, we examine the role that remittances sent two years prior have on reporting of remittances in the current year. The

<sup>&</sup>lt;sup>8</sup> See McKenzie, et al (2007) for a discussion of the difficulties in estimating the effect of remittances on desired outcomes. Also see Adams (2006) and Rapoport et al (2006) for reviews of the kinds of estimations typically conducted in the literature on remittances.

<sup>&</sup>lt;sup>9</sup> The eight different categories were: direct credit to NRI accounts, demand drafts, internet based money transfer service, formal money transfer services, family member cash transfer, friends cash transfer, international debit or credit cards, other means.

lagged remittance amount appears to be negatively related to the self-reported remittance amount in the current year. This is additional evidence for strong mean-reversion in the data on remittances. These results indicate that survey respondents are using multiple reference points when responding to the survey questionnaire regarding remittances, much like the Rashomon effect's subjectivity of perception on recollection.

Finally, we explore how the non-classical measurement error in remittance data can adversely affect ordinary least squares regression results. We use the self-reported information on remittances to explain household expenditures on food. Acknowledging the multiple biases present in a simple cross-section regression, we employ an instrumental variables approach and compare the results once self-reported remittances are properly instrumented. Our results are troubling; the instrumental variables approach may solve problems of endogeneity but not the problem of non-classical measurement error. Our IV estimates is upwardly biased from the true slope parameter and is of no greater value than the initial OLS estimate. While we do not have a definitive solution to this problem, we do offer some suggestions for further research in the discussion. We also note that our results may indicate that it is preferable to ask detailed questions about the different channels and the amounts remitted via these channels as opposed to simply asking a general question about remittances.

The next section discusses previous remittance research. The following section provides information on the data collection and characteristics of the data set. In Section 4 we describe the nature and determinants of the measurement error contained in the remittance data. We discuss the implications of measurement error in an ordinary least squares regression and in an instrumental variables regression in Section 5. In Section 6 we explore the future research efforts in this area and conclude in Section 7.

#### II. Previous Literature

There have been very few studies examining the discrepancies in actual and reported remittance flows. Existing studies have compared differences in Balance of Payment data and aggregated household survey amounts. For instance, in Kyrgyz Republic, Andriasik (2005) finds that the remittances recorded in the Balance of Payments figures are twice as large as the household survey remittance aggregate. Shonkwiler, et al (2011) find qualitatively similar results in Armenia and propose an econometric solution for underreporting. Acosta et al (2006) find similar results for 11 countries in Latin America. Our study differs from these previous ones in that we have actual microeconomic data between self-reported household surveys and administrative measures of flows to the NRI accounts.

In this regard, our research fits into the existing measurement error literature on self-reported salary and wage income. Several authors have investigated the nature of measurement error in US wage and salary data. Duncan and Hill (1985), Bound and coauthors (1991;1994;2001), Pischke (1995) and Moffit and Gottschalk (2002) have found that the self-reported data tends to contain non-classical measurement error with mean reversion. Akee (2011) found similar results in a developing country context; however, the degree of measurement error was far worse than that in the US. Antman and McKenzie

(2007) examine the implications of measurement error on measures of earnings mobility. These studies are unique in that they have a generally accepted check on self-reported information which is generally unavailable in standard data sets. Black et al (2000) provide bounding parameters when data exhibit non-classical measurement errors.

## III. Data Set Collection and Description

#### a. Data Set Creation

We obtained the data for this analysis from the Reserve Bank of India. The research division commissioned a survey of households that had an operating NRI account in the four states of Gujarat, Kerala, Maharashtra and Punjab. Therefore, the population studied here are households with an operating (any transactions in the last two years) NRI account in these four Indian states. The household survey was intended to provide information on the amounts, uses and characteristics of migrants in NRI account-holding households in four Indian states that receive the highest volume of remittances. A small subset of this household survey was linked to RBI administrative data on NRI account transactions. Our analysis in this paper is conducted on the subset of matched data.

A survey firm, hired by the RBI, conducted the household survey from December 2009-July 2010. The RBI provided the survey firm with approximately 4000 names and addresses of randomly selected NRI account holder households. Enumerators contacted households by phone and by going to the address directly. The final household survey contained 2756 observations for a response rate of almost 70%.

The household survey contained several modules with questions on household demographics, consumption, asset ownership and liabilities, migrant information, uses of remittances, pattern of debits and credits to NRI accounts, and finally a section on the modes of sending remittances. The official, administrative data from the RBI provided information on the number of credits and debits to the NRI accounts in the previous two years as well as the total amount credited to the account in the previous year. The RBI provided account data for approximately 50 accounts from each of the 4 different states; these accounts were merged according to address information. The characteristics of the matched and non-matched sample do not differ significantly with respect to education levels, household income levels, or marital status. However, the matched sub-sample is about 5 years older than the non-matched sample and is 10 percentage points more likely to be male.

#### **b.** Data Description

We present a description of the data in Table 1. Our results indicate that on average households receive 600,000 rupees remittances annually via the non-resident accounts, while these same households only report 290,000 rupees. There is a large under reporting of the amount of remittances received from household migrants. The sample size is reduced to 209

observations for the remained of our analysis as this omits observations with missing information, etc. <sup>10</sup>

The average age of survey respondents is about 51 years old and about 60% are males. Over 90% of the survey respondents are married and are relatively highly educated, with over 75% of the sample having tenth standard education or higher. Finally, we report the distribution of households among the four different survey areas: 30% of our sample comes from Gujarat, 20% from Kerala, Maharashtra has almost 29% and Punjab 22%. Overall these characteristics indicate that our sample is comprised of well-educated households. Comparing these results to the National Sample Survey of Indian 64th Round (NSS), our average household is highly educated; the average household head in the NSS has a fifth standard education or less. The age and marital rates are approximately similar, however. Our sample, based on households that have at least one NRI account, is drawn from the upper end of the education distribution.

In Figure 1 we present the kernel density figures for the number of times there were actual credits to NRI accounts and the survey reported amount of credits. The dashed line represents the number of times a migrant sends credits to an NRI account in the past year; the modal number of times is approximately 2 to 3 times. On the other hand, the matched household survey respondent indicates that the account receives a modal amount of 12 deposits per year. Figure 2 provides a similar analysis for the year 2007. While the difference is not as dramatic as in the previous figure, we find that the results are qualitatively very similar across the two time periods.

In addition to the measurement error in the credits to the NRI accounts, we also find evidence for measurement error in debits to these same accounts. The next two figures provide the number of actual and reported debits to the NRI accounts over the course of 2008 and 2007. Once again there is an over reporting of the number of transactions in the NRI accounts as compared to the actual number of debits in the year. It is not clear whether this is simply recall bias or some sort of skewing of the answer in order to impress the surveyor. Nevertheless, it is clear that there is over reporting for both the number of credits and debits to the NRI accounts.

We report the difference in the actual remittance amounts credited by migrants abroad to NRI accounts in 2008 relative to the survey reported amounts. Figure 5 provides the two figures for the self-reported and administrative bank data. The raw data here indicates that there are more people reporting relatively low amounts of remittances than actually receive low remittances. On the other hand, very few people tend to report that they are receiving

<sup>&</sup>lt;sup>10</sup> It is important to note that previous studies on measurement error in survey responses typically have similar small sample sizes. For instance, Bound and Krueger (1991) have just 444 observations in their Current Population Survey – Social Security match dataset for their research on measurement error in wage reporting; Pischke (1995) has just 234 observations in his panel data using the Panel Study of Income Dynamics (PSID) data and 351 (437) observations in his 1982 (1986) cross-section.

large remittances. The number of respondents reporting large amounts of remittances is considerably smaller than the number actually receiving these large remittance amounts. This figure points to the conclusion that households are generally under reporting the size of the annual amount deposited to NRI accounts.

Overall, the figures indicate that measurement error is present in several categories of variables and over different years. Individual survey respondents tend to underreport the amount of the actual remittances received via NRI accounts. Additionally, they tend to over report the frequency of remittances sent from abroad.

#### IV. Characterization of the Measurement Error

It is clear from the raw data that there is a large discrepancy between the actual and reported credits to NRI accounts. We explicitly compute the error amount as the following:

(1) Error = Reported Amount - Actual Amount

The calculation above is a standard method in the measurement error literature (Bound and Krueger, 1991).

#### a. Graphical Depiction of the Remittance Measurement Error

It is useful to get a sense of how the measurement error in our data relates to the actual value of the remittance amount. We graph this relationship in Figure 6. The downward slope of the data indicates that the measurement error in the reporting of deposits to NRI accounts is non-classical. If the measurement error were classical (white noise) in nature, the differences would be equally distributed around a horizontal value of 0. In this case, we find that the measurement error is more negative as the actual value of remittances increases. This negative relationship is comparable to previous findings with regard to earnings and wages for the US and other countries. To our knowledge, this is the first documentation of the existence of non-classical measurement error in remittance reporting.

Figure 8 provides a simple histogram of the distribution of the errors in self-reported and administrative reports of credits to NRI accounts. It is encouraging that much of the data is centered on zero, but there are significant outliers in both directions. The left hand tail is larger than the right which is indicative of the fact that the self-reported remittances tends to be lower than actual remittances; this fact was already noted in Table 1.

#### b. Correlation Coefficients and Reliability Ratios of the Measurement Error

Now, we investigate the relationship between the measurement error and the actual and self-reported remittances numerically. We provide correlation coefficients and reliability ratios for the self-reported deposits to NRI accounts. The reliability ratio gives the signal to noise ratio for the variance of the actual remittances.

The first row in panel A indicates that the actual remittance amounts are positively correlated with the self-reported remittance amounts. The second row indicates that the administrative data and the measurement error are negatively related; this is the same result found graphically in Figure 6. Bound and Krueger termed this effect "mean reversion". Finally, the third row indicates that the self-reported remittance amount and measurement error is positively related. This panel provides strong evidence that the measurement error in the remittances sent to NRI accounts is non-classical.

The second panel of Table 2 provides the reliability ratios of the self-reported information on deposits to the NRI accounts. The two results differ according to whether the data contains classical or non-classical measurement error. If we assume that the measurement error is classical, then the reliability of the measure is approximately 0.52. However, as noted previously, we have strong evidence that the measurement error is not classical. Therefore, the accurate depiction of the reliability ratio is given, instead, by the next value: 0.703. This value indicates that almost 70% of the variance in the self-reported remittance amount can be attributed to variance in the actual remittances received. Overall this appears to be good news.

#### c. Determinants of Measurement Error Regression Analysis

It is also useful to identify whether there are certain household characteristics that are associated with the reporting error in remittances. We regress the error amount defined in equation 1 on a vector of household characteristics.

(2) 
$$Error_i = \beta_0 + X'\beta_1 + \varepsilon_i$$

In the equation above, the vector X contains household characteristics such as education, gender and age of the survey respondent.

Additionally, in a separate regression, we add a measure of the total amount of remittances received by the household in the year before last. Inclusion of this additional variable helps to indicate whether shocks to previous years' remittance amounts affects reporting in the current year. It is possible that individuals base their reporting amounts on a long-run average of remittances. Similar to the permanent income hypothesis, people may have a long-run view of remittances and report an average (across years) amount. Slight deviations and transitory changes are not well reported (Pischke, 1995).

The regression results are provided in Table 3. We find that a person from the lower end of the education distribution is more likely to under report the remittances received in an NRI account. Given how we defined measurement error in equation 1, for a given level of actual remittances an individual with lower education levels (relative to the high education omitted category) is more likely to under report the remittances that they receive as compared to the actual amount. These results are statistically significant at the 5% level. Additionally, if we restrict the comparison to a binary one with an individual obtaining less

than 12th Standard or not, the coefficient continues to be statistically significant and negative. This indicates that, relative to the high skilled respondents, these lower skilled individuals are more likely to under report the remittance amount. The other characteristics such as age, the gender indicator variable and marital status do not appear to be related to the reporting error.

The second column of Table 3 repeats the regression from the first column with one additional control variable – the amount credited to the NRI account in the previous year which is provided from the administrative data. Overall, the main results appear to be quite similar with lower levels of education associated with lower reporting of remittances on the surveys. Interestingly, we find that the total amount credited to NRI accounts in the previous year has a negative effect on the size of the reporting error. These results mean that a household that received a high amount of remittances in the previous year is more likely to underreport their remittances in the current year. This suggests the potential for some intertemporal mean regression; in fact, an individual household could be reporting some long-run average of NRI account transmittals. Fluctuations may not be well reported. There is much room for further investigation of the dynamic effects of fluctuations in migrant remittances on current levels of remittance reporting.

## V. Potential Bias in Estimation Results when Remittances Measures contain Measurement Error

It is a well-known fact that classical measurement error leads to attenuation bias in ordinary least squares regressions. The use of instrumental variables in these cases will result in an estimated coefficient that is significantly larger than the OLS coefficient. In the presence of non-classical measurement error, however, an instrumental variables approach may be upwardly biased.

Assume that we are concerned with the following cross-section analysis:

(3) 
$$Y_i = \beta' X_i + u_i$$

Our interest is in estimating the effect of X, which can be thought of as remittances received by the individual household i on a measure of investment or consumption, Y.

In the presence of classical measurement error in the X variable, ordinary least squares estimates of the slope coefficient tends to be biased towards zero. Assume that the measurement error in the right hand side variable is determined in the following manner:

$$(4) X = X^* + \varepsilon$$

<sup>&</sup>lt;sup>11</sup> It is important to note that the sample size decreases here because not all households had remittances sent to their NRI accounts in 2007.

where X is the observed value of remittances received self-reported in a household survey, for instance. The variable  $X^*$  is the true (unobserved) measure of remittances and  $\varepsilon$  is the measurement error. In this case, ordinary least squares will produce the following slope estimate:

$$p \lim \hat{\beta}_{OLS} = \beta \left( \frac{\sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_{\varepsilon}^2} \right) < \beta$$

The estimated coefficient will be downward biased away from the true population parameter as long as  $\sigma_{\varepsilon}^2 > 0$ . The use of instrumental variables will solve this problem. If a valid instrument Z exists and satisfies the conditions that Cov(Z,u) = 0 along with the assumption that the measurement is classical  $Cov(X^*,\varepsilon) = 0$ , then the estimated coefficient is unbiased:

$$p \lim_{\beta_{IV}} \hat{\beta}_{IV} = \left(\frac{\beta \sigma_{X^*,Z} + \sigma_{Z,u}}{\sigma_{X^*,Z} + \sigma_{X^*,E}}\right) = \beta \left(\frac{\sigma_{X^*,Z}}{\sigma_{X^*,Z}}\right) = \beta$$

However, in addition to measurement error, the ordinary least squares regression in this particular scenario also suffers from endogeneity bias. This is especially true in the case of using remittances to explain household consumption or investment: poorer households may receive more remittances to help with household consumption or poorer households may receive less remittances since migrants from these households may not earn enough abroad to send large sums home. In either case, there is a well-known endogeneity problem in cross-section regressions. In this case, we would have an additional term to contend with for the OLS coefficient:

$$p \lim \hat{\beta}_{OLS} = \beta \left( \frac{\sigma_{X^*}^2 + \sigma_{X^*,\varepsilon}}{\sigma_{X^*}^2 + 2\sigma_{X^*,\varepsilon}^2 + \sigma_{\varepsilon}^2} \right) + \left( \frac{\sigma_{X^*,u}}{\sigma_{X^*}^2 + 2\sigma_{X^*,\varepsilon}^2 + \sigma_{\varepsilon}^2} \right) = \beta \left( \frac{\sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_{\varepsilon}^2} \right) + \left( \frac{\sigma_{X^*,u}}{\sigma_{X^*}^2 + \sigma_{\varepsilon}^2} \right) < or > \beta$$

In addition to the attenuation bias term, we would have an additional term on the right hand side which is the endogeneity bias. Note that the term,  $2\sigma_{X^*,\varepsilon}$  is zero by the classical measurement error assumption. The endogeneity bias amount may be either positive or negatively biased depending upon the relationship between X and the error term u. In this situation, an instrumental variable would also solve the endogeneity bias as well as the attenuation bias induced by classical measurement error.

$$p \lim \hat{\beta}_{IV} = \frac{Cov(Z,Y)}{Cov(X,Z)} = \frac{Cov(Z,\beta X^* + u)}{Cov(X^* + \varepsilon,Z)} = \left(\frac{\beta \sigma_{X^*,Z} + \sigma_{Z,u}}{\sigma_{X^*,Z} + \sigma_{X^*,E}}\right) = \beta \left(\frac{\sigma_{X^*,Z}}{\sigma_{X^*,Z}}\right) = \beta$$

In this case, the instrumental variable Z, if valid, ensures that the resulting coefficient is an unbiased estimate of the true population parameter.

If the assumption of classical measurement error does not hold, however, then instrumental variables can only serve to eliminate the endogeneity bias and not the bias due to measurement error. The estimated coefficient from the instrumental variables regression, while free of endogeneity bias, may still have an upward bias due to the presence of non-classical measurement error.

$$p \lim_{\beta_{IV}} = \frac{Cov(Z,Y)}{Cov(X,Z)} = \frac{Cov(Z,\beta X^* + u)}{Cov(X^* + \varepsilon,Z)} = \left(\frac{\beta \sigma_{X^*,Z} + \sigma_{Z,u}}{\sigma_{X^*,Z} + \sigma_{X^*,\varepsilon}}\right) = \beta \left(\frac{\sigma_{X^*,Z}}{\sigma_{X^*,Z} + \sigma_{X^*,\varepsilon}}\right) > \beta$$

In the equation above, the instrument ensures that the term  $\sigma_{Z,u}$  is equal to zero. However, now in this situation,  $\sigma_{X^*,\varepsilon} \neq 0$ , and it is in fact negative; this is the "mean reversion" we found in our data and others have found in other wage and earnings data sets. The instrumental variables estimate of the effect of X on Y is biased upwards.

Overall these results are somewhat unsettling. In the presence of non-classical measurement error, the instrument variable method provides an upwardly biased coefficient. At best, we can consider the instrumental variables coefficient to be an upper bound on the true size of the population parameter. However, because we are dealing with two issues in the cross-section situation (endogeneity bias and attenuation bias) we can not assume that the OLS coefficient is a lower bound of the true coefficient. In fact, the OLS coefficient may be upwardly biased as well when the measurement error is not classical.

$$p \lim_{\beta_{OLS}} = \beta \left( \frac{\sigma_{X^*}^2 + \sigma_{X^*, \varepsilon}}{\sigma_{X^*}^2 + 2\sigma_{X^*, \varepsilon} + \sigma_{\varepsilon}^2} \right) + \left( \frac{\sigma_{X^*, u}}{\sigma_{X^*}^2 + 2\sigma_{X^*, \varepsilon} + \sigma_{\varepsilon}^2} \right) < or > \beta$$

The estimated OLS coefficient in this case may be either larger or smaller than the true population coefficient depending on the signs of the covariance of  $X^*$  and u and the magnitude of the mean reversion,  $\sigma_{X^*,\varepsilon}$ . In practical terms this means that the actual coefficient may lie outside the interval of the OLS and IV estimated coefficients.

As an example, we provide a consumption regression in Table 4. We regress the log of household food consumption on the amount of annual remittances (in 100,000s rupees) reported by the household. In the first column we report the ordinary least squares regression results. The results indicate that an increase of 100,000 rupees in annual remittances to the NRI accounts will increase the monthly household food expenditure by 2.6%. Since we know that this ordinary least squares regressions suffers from both endogeneity and measurement error bias, we attempt to correct for this by using an instrumental variables approach.

In columns two and three we present the first and second stage regressions respectively. In the first stage, we use the per capita GDP growth in 2007 in the destination country as an instrument for the amount of remittances deposited to the NRI accounts by migrants already located in the destinations. The instrument has a positive effect on remittances, as expected and is not a weak instrument with an F-statistic of 10.34. In column 3, we present the second stage regressions with the log of monthly food expenditures as the outcome variable. The instrumented coefficient on the amount of remittances received is more than three times larger than the OLS coefficient in column 1. The results indicate that an increase of 100,000 rupees in annual remittances to the NRI accounts will increase the monthly household food expenditure by 10.6%. These are large amounts and would indicate a big effect of remittances on household food consumption if correct. However, as noted previously this IV coefficient suffers from biases due to the non-classical measurement error in the data.

Finally, in column 4, we present the ordinary least squares regressions using the actual amount each household received in NRI deposits over the course of 2008. As mentioned previously the actual amount remitted is about double the reported amount of remittances. While the OLS regression is free from measurement error, it still contains endogeneity bias. We were unable to identify a valid instrument for the remittance variable in this case. Therefore, we are unable to determine the true population parameter; however, we are able to sign the likely direction of the endogenity bias. If we assume that the most important omitted variable is household preferences for food it is likely that this is positively correlated with remittances. Households with a preference for large expenditures on food are also more likely to request more remittances. We would then conclude that our estimated coefficient in column 4 would have a positive bias. Ultimately, the effect of remittances on household food expenditures would appear to be close to zero. This is not incredibly surprising as NRI account households tend to be relatively wealthy and probably not in need of remittances for household consumption needs or funds for household food consumption may be remitted primarily through other channels.

## VI. Discussion - Implications for Research

The results for measurement error in the flow of remittances via NRI bank accounts are not encouraging. It appears that the data exhibits strong non-classical measurement error that exacerbates biases in regression coefficients. Other channels of remittance flows may also exhibit this same kind of measurement error; so far this has gone undocumented.

<sup>&</sup>lt;sup>12</sup> McKenzie et al (2006) discuss the inherent problems in using non-experimental data in estimating the effects of migration on wages. In a unique setting with an experimental migration setting, they find overestimates of the true effects of migration on wages when using instrumental variables. They attribute the overestimate to the validity of their instruments. Our analysis here, while focusing on remittances, may suggest that their overestimates from instrumental variables estimation may also be partially attributable to non-classical measurement error in the data.

Uncovering the causes of reporting error would be useful in the development of better survey questions.

In this section we examine additional determinants of measurement error. We used the head of household characteristics in Table 3 to explain measurement error and found that the lower educated heads of households tended to underreport the amount of remittances sent via NRI accounts. We now augment that analysis with characteristics of the migrant himself.

The regression results are presented in Table 5. The first column adds both state dummy variables as well as whether the survey respondent is the head of household. The coefficient on that variable is negative and statistically significant which indicates that household heads tend to underreport the receipt of remittances. The literature on measurement error identifies recall bias and social desirability reporting as two potential causes of mis-reporting (Bound, et al, 2001). It is difficult to expect that the household head (relative to other household members) would have less information about the actual NRI deposits made from a migrant abroad. On the other hand, it might be more likely that a head of household is making a more strategic decision in answering the survey question on remittances. Instead of trying to impress the survey enumerator as suggested under social desirability reporting, the respondent may simply be trying to disguise or hide the remittance amounts. As mentioned earlier the survey was conducted on behalf of the Reserve Bank of India and there may have been concern that the information may be used to implement some form of taxation on these assets. Overall, the results indicate that heads of households are systematically underreporting NRI account deposits.

The second column includes information on the migrant, as reported by the survey respondent. Age, education level and marital status of the migrant do not appear to have an effect on errors in remittance reports. However, the gender of the migrant does cause the survey respondent to underreport NRI account deposits. When the migrant who remits is a male, the respondent is likely to underreport remittance amounts by an additional Rs 420,000. This may simply be explained by the fact that male migrants tend to send much more than female migrants and tend to be the majority of remittance senders in this data.

Finally, in the third column, we add in whether or not the account is jointly administered by the survey respondent with the migrant abroad. The alternative is that the account is private and only accessed by the migrant himself. This information, which should have clear indication for the amount of information that the survey respondent has regarding the NRI account. A jointly held account will mean that the individual has much more accurate information regarding the number and amounts of credits from abroad. As expected, this variable has a strong and statistically significant effect on the reporting NRI credits. A positive coefficient on this variable indicates that the survey respondents who jointly administered the NRI account with the migrant tended to report more accurately the NRI deposits; in this data the average measurement error is strongly negative, therefore any positive coefficient results in a more accurate reporting. This result is particularly encouraging. Including an additional survey question about whether the individual jointly

administers this NRI account may provide researchers with additional information on the quality of responses.

#### VII. Conclusion

We find that the reporting of a specific remittance flow, deposits to NRI accounts, suffers from measurement error. Individuals report a much higher frequency of deposits than actually occurs as documented from bank transactions data. Additionally, the total annual amount credited to NRI accounts is much larger than self-reported total remittances. These differences in remittance flows are not merely white noise; it is negatively related to the actual amount of remittances sent.

In our work we have emphasized the implications that this non-classical measurement error has for both macroeconomic and microeconomic research. At the larger macroeconomic level, the discrepancies once aggregated over different remittance channels may partially explain the existing discrepancies in official macroeconomic data and aggregated nationally-representative household surveys of remittance amounts in India. In microeconomic household-level data, it is clear that there may be severe biases in regression coefficients which use remittances as a causal variable when it contains non-classical measurement error. Unfortunately, we also find that instrumenting for the error may also produce biased results in the presence of non-classical measurement error.

Our research suggests that there is considerable room for improvement in the collection of remittance data and survey questions. Experimentally testing and validating different forms of remittance survey questions would be an important step in assessing the usefulness of remittance survey questions (see, for instance De Mel, et al 2007). We have shown that the addition of a single survey question which indicates the survey respondent's knowledge and legal access to the NRI accounts greatly improves reporting accuracy.

Finally the results point to the need to more carefully check the accuracy of the international remittance flows. This would require comparing aggregated data from nationally representative household surveys with improved questionnaire design (as suggested in this paper) with Balance of Payment data on the one hand, as well as comparing bilateral outflows and inflows of remittances from sending and receiving countries on the other. The wide discrepancies in the Indian case, to take the example for this paper, could be both because of inaccuracies in the household survey as well as mis-classification of the Balance of Payment data with some fraction of reported remittances being disguised capital flows (and hence likely to be less stable) rather than current account flows for family maintenance. While the precise explanation needs further research, the implications of these inaccuracies, as our paper demonstrates, are substantial.

### References

- Adams, Richard. 1998. "Remittances, investment and rural asset accumulation in Pakistan." Economic Development and Cultural Change. V. 47, pp. 155-173.
- Adams, Richard. 2006. "International remittances and the household: analysis and review of global evidence." Journal of African Economies. V. 15 no 2, pp. 396-425.
- Adams, Richard and J. Page. 2005. "Do International Migration and Remittances reduce poverty in developing countries?" World Development. V. 33 no 10, pp. 1645 1669.
- Adams, Richard and A. Cuecuecha. 2010. "Remittances, Household Expenditure and Investment in Guatemala." World Development. V 38 no 11, pp. 1626-1641.
- Akee, R. 2011. "Errors in self-reported earnings: The role of previous earnings volatility and individual characteristics. Journal of Development Economics, forthcoming.
- Angrist, J. and A. Krueger. 1999. "Empirical Strategies in Labor Economics." in Handbook of Labor Economics. O. Ashenfelter and D. Card, eds. Amsterdam: North-Holland. V. 3A, 1277-1366.
- Antman, Francisca and David J. McKenzie. 2007. "Earnings Mobility and Measurement Error: A Pseudo-Panel Approach." Economic Development and Cultural Change, 56(October): 125-161.
- Black, D. and M. Berger and F. Scott. 2000. "Bounding Parameter Estimates with Nonclassical Measurement Error." Journal of the American Statistical Association, 95(451):739-748.
- Bound, John and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, 9(1): 1-24.
- Bound, J. and Charles Brown, Greg J. Duncan and Willard L. Rodgers. 1994. "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data." Journal of Labor Economics, 12(3): 345-368.
- Bound, J. and Charles Brown and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In The Handbook of Econometrics, Volume 5. eds. James J. Heckman and Edward Leamer, 3705-3843. New York: Elsevier Science. Cappellari, Lorenzo. 1999. "Minimum Distance Estimation of Covariance Structures." Fifth UK Meeting of STATA Users, Royal Statistical Society, London, UK.
- Brown, R. and E. Jimenez. 2008. "Estimating the Net Effect of Migration and Remittances on Poverty and Inequality: Comparison of Fiji and Tonga." Journal of International Development V 20, pp. 547-571.
- Cox Edwards, A. and M. Ureta (2003): International migration, remittances and schooling: evidence from El Salvador, *Journal of Development Economics*, 72, 2: 429-61.
- de la Brière, B., E. Sadoulet, A. de Janvry and S. Lambert. 2002. "The roles of destination, gender, and household composition in explaining remittances: An analysis for the Dominican Sierra." *Journal of Development Economics*, 68, 2: 309-28.
- De Mel, Suresh and David McKenzie and Christopher Woodruff. 2007. "Measuring Microenterprise Profits: Must we ask how the sausage is made?" Journal of Development Economics, doi: 10.1016/j.jdeveco.2008.01.007.

- Duncan, Greg J. and Daniel H. Hill. 1985. "An Investigation of the Extent and Consequences of Measurement Error in Labor-economic Survey Data." Journal of Labor Economics, 3,(4): 508-532.
- El- Qorchi, Mohammed, Samuel Munzele Maimbo, and John Wilson, 2003, "Informal Funds Transfer Systems: An Analysis of the Informal Hawala System." IMF Occasional Paper No. 222 (Washington: International Monetary Fund).
- Funkhouser, E. (1995): Remittances from international migration: a comparison of El Salvador and Nicaragua, *Review of Economics and Statistics*, 77, 1: 137-46.
- Lucas, R and O. Stark. 1988. "Migration, remittances, and the family." Economic Development and Cultural Change. V 36 no 3, pp. 465 481.
- McKenzie, David; Gibson John; Stillman, Steven. 2006. "How Important is Selection? Experimental versus Non-Experimental Measures of the Income Gains from Migration." World Bank Policy Research Working Paper 3906.
- McKenzie, David and M. Sasin. 2007. Migration, Remittances, Poverty and Human Capital: Conceptual and Empirical Challenges. World Bank Policy Working Paper No. 4272. Washington, DC: World Bank.
- Moffit, Robert and Peter Gottschalk. 2002. "Trends in the Transitory Variance of Earnings in the United States." The Economic Journal. 112(March): 68-73.
- Pischke, Jorn-Steffen. 1995. "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." Journal of Business and Economic Statistics. 13(3): 305-314.
- Rapoport, Hilel and F. Docquier. 2006. "The economics of migrants' remittances. In: Kolm, S. and J. Mercier Ythier (Eds.), Handbook of the Economics of Giving, Altruism and Reciprocity. North Holland, pp. 1135-1198.
- Reserve Bank of India, 2006a, "Report of the Working Group on Cost of NRI Remittances," May. Available at: http://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/72983.pdf
- Reserve Bank of India, 2006b, "Remittances from Overseas Indians: A Study of Methods of Transmission, Cost, and Time," *RBI Bulletin* November 2006, Annex 3.
- Reserve Bank of India. 2008. "India's Balance of Payments Developments during Fourth Quarter of 2007-08" Unpublished Document.
- Shonkwiler, J. Scott and David Grigorian and Tigran Melkonyan. 2011. "Controlling for the underreporting of remittances." Applied Economics, 43(30), 4817-4826.
- Yang, Dean. 2008. "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks." *Economic Journal*, 118(528): 591-630.

## Appendix Table 1.

## **US-India Remittance Flows**

	2006-07	2007-08	2008-09
CBO	2.475	2.975	2.975
RBI	10.02	14.24	13.79

Source: Congressional Budget Office Migrants' Remittances and Related Economic Flows February 20111, Exhibit 4; RBI Bulletin, Table 6, p. 787, April 2010.

Table 1: Means and Standard Deviations

St. Dev.
1223161
507149.7
3.2 1179510
4 1065962
6 335627
8 14.823
0.498
0.281
0.281
0.351
0.501
0.441
0.458
0.402
0.453
0.412

Note: Sample size is 209 observations

Table 2: Simple Correlations and Reliability Ratios for Administrative, Reported Earnings Data and Reporting Errors

A. Correlation Coefficients	
Correlation ( Administrative Remittances, Self-Reported Remittances)	0.291
Correlation (Administrative Remittances, Measurement Error)	-0.911
Correlation (Self-Reported Remittances, Measurement Error)	0.127
B. Reliability Ratios	
Reliability Ratio for Classical Measurement Error Reliability Ratio = True Measure Variance / (Error Variance + True Measure Variance)	
Reliability Ratio =	0.518
Reliability Ratio for Non-Classical Measurement Error Reliability Ratio = Covariance(Inadmin, Inreport) / Variance (Inreport)	
Reliability Ratio =	0.703

Note: Sample size = 209

Table 3: Regression of Differences in Self-Reported and Administrative Records of Remittances for 2008

	(1)	(2)
VARIABLES	Difference in Rupees (000's)	Difference in Rupees (000's)
Education Variables		
Less Than 5th Standard	55.535	129.960
	(199.903)	(237.438)
Fifth Standard to Ninth Standard	-251.187	-400.430**
	(219.467)	(182.340)
Tenth to Twelfth Standard	-399.266**	-345.946**
	(194.460)	(154.130)
Greater Than 12th Standard	Reference Category	Reference Category
Demographic Variables		
Age	-0.352	6.265
	(5.913)	(5.114)
Male Indicator Variable	41.693	-105.025
	(167.604)	(168.011)
Married Indicator Variable	-67.825	30.631
	(184.544)	(233.825)
Total Amount Credited to NRI Accounts in 2007 in		-0.389***
(000's)		(0.033)
Constant	-18.492	-45.029
	(328.738)	(388.609)
Observations	209	143
R-squared	0.0267	0.539

Robust standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

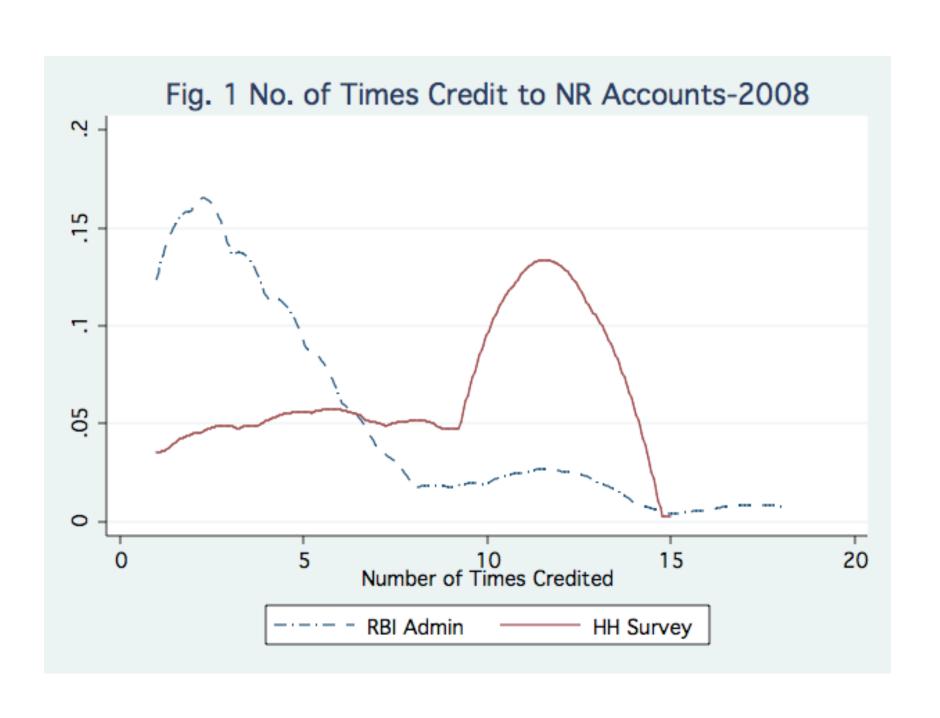
Table 4: Average Household Monthly Food Consumption Regression (3) (4) OLS IV IV OLS Amount of Remittances Self-Log of Average Log of Average Food Food Log of Average Food Reported in Ruppes **VARIABLES Expenditures** (100,000's)Expenditures Expenditures Survey Reported Annual Amount of Deposits to 0.026\*\*\* 0.106\*\* NRI Accounts in 100,000 rupees (800.0)(0.042)Administrative Annual Amount of Deposits to 0.002 NRI Accounts in 100,000 rupees (0.003)Instrumental Variable GDP Growth in Destination Country in 2007 0.769\*\*\* (0.232)Education Variables -0.399\*\*\* -0.467\*\*\* Less Than 5th Standard -3.309\*\* -0.175(0.152)(1.356)(0.221)(0.155)-2.728\*\* Fifth Standard to Ninth Standard -0.1420.046 -0.207(0.134)(1.194)(0.192)(0.137)Tenth to Twelfth Standard 0.075 -1.724\*\* 0.234\* 0.017 (0.093)(0.821)(0.140)(0.094)Greater Than 12th Standard reference reference reference reference Demographic Variables 0.002 0.036 -0.001 0.002 Age (0.004)(0.003)(0.028)(0.003)Male Indicator Variable 0.016 1.084 -0.039 0.034 (0.087)(0.788)(0.112)(0.089)Married Indicator Variable 0.126 0.310 0.062 0.142 (0.140)(1.262)(0.177)(0.144)Urban Resident 0.162\*\* 2.657\*\*\* -0.037 0.215\*\*\* (0.081)(0.710)(0.143)(0.083)Asset Variables -0.669 0.028 Own this residence? -0.1427.128 (0.525)(4.677)(0.704)(0.536)Number of livestock owned 0.056 0.078 0.049 -0.251(0.056)(0.502)(0.071)(0.058)0.192\*\*\* 0.179\*\*\* Number of vehicles owned -0.3840.234\*\*\* (0.027)(0.239)(0.039)(0.027)Constant 8.603\*\*\* -7.585 9.001\*\*\* 8.476\*\*\* (0.570)(5.151)(0.736)(0.584)Observations 199 199 199 199 R-squared 0.361 0.187 0.0147 0.326

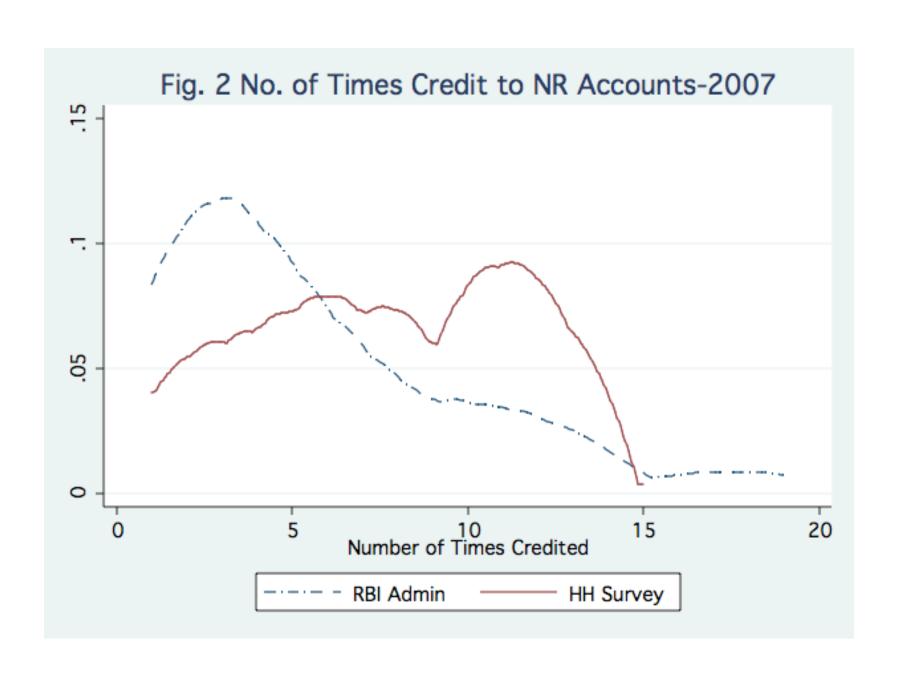
Standard errors in parentheses

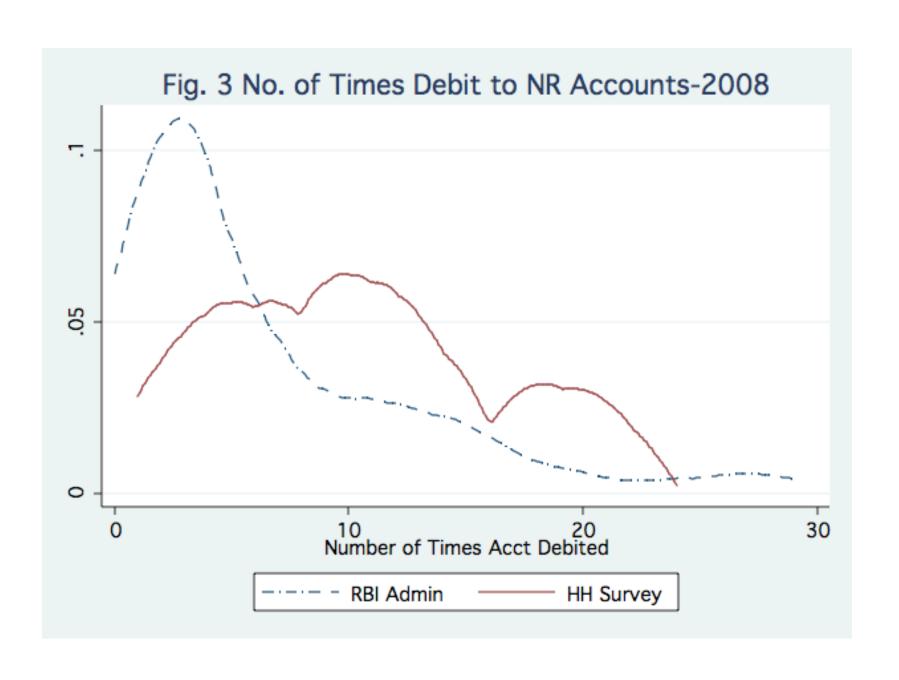
<sup>\*\*\*</sup> p<0.01, \*\* p<0.05, \* p<0.1

	(1)	(2)	(3)
	Measurement	Measurement	Measurement
WARLARI EG	Error in (000)	Error in (000)	Error in (000)
VARIABLES	Rs.	Rs.	Rs.
NRI Account Jointly Administered?			757.418**
			(378.971)
Current Age of Migrant		10.985	12.399
		(8.913)	(8.056)
Education of Migrant Prior to Migration		-14.152	-19.033
		(41.681)	(40.903)
Gender of Migrant		-420.362**	-472.378**
		(185.073)	(184.534)
Current Marital Status of Migrant		-41.155	-162.486
		(287.666)	(311.724)
Resondent is Head of Household	-420.116**	-460.105**	-605.472***
	(162.164)	(177.534)	(199.462)
Less Than 5th Standard	-187.545	-255.080	-280.006
	(215.402)	(224.786)	(230.222)
Fifth Standard to Ninth Standard	-178.831	-243.220	-128.597
	(236.928)	(244.316)	(261.047)
Tenth to Twelfth Standard	-274.716	-280.175	-242.202
	(186.968)	(190.799)	(185.499)
Greater Than 12th Standard	Reference	Reference	Reference
Age	1.185	0.375	-1.886
	(7.155)	(7.404)	(7.623)
Male Indicator Variable	295.953	329.228	425.618*
	(214.561)	(206.416)	(216.448)
Married Indicator Variable	-202.527	-258.741	-309.577
	(206.924)	(219.349)	(230.004)
Gujarat	62.374	11.078	-87.413
-	(317.283)	(342.676)	(354.655)
Kerala	804.409**	845.136**	733.004**
	(309.888)	(328.536)	(341.569)
Maharashtra	474.692	478.744	563.375*
	(294.610)	(304.085)	(307.401)
Punjab	Reference	Reference	Reference
Constant	-89.094	144.117	-151.925
	(530.641)	(910.673)	(830.761)
Observations	209	209	209
R-squared	0.0922	0.106	0.143

Robust standard errors in parentheses
\*\*\* p<0.01, \*\*\* p<0.05, \* p<0.1







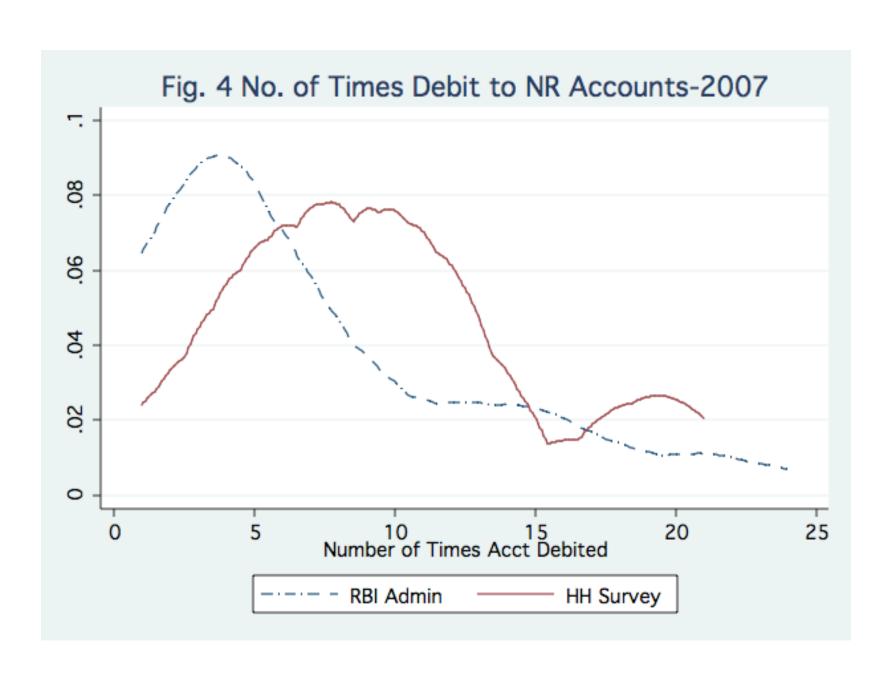


Fig. 5 Amt of Credit to NR Accounts-2008

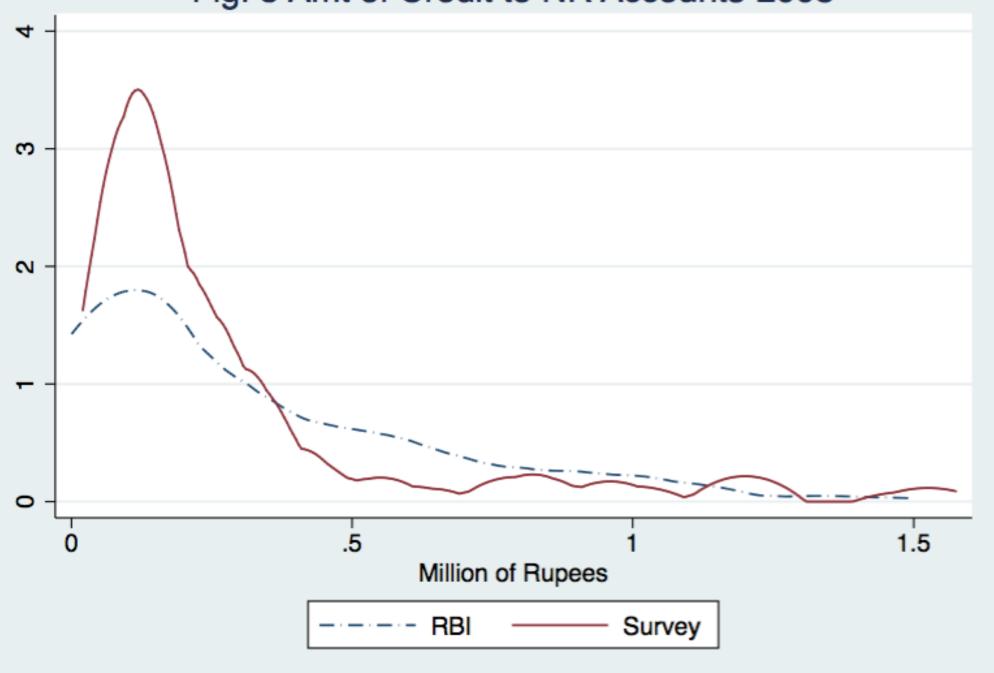


Fig. 6 Diff in Reported and Admin Credit Amt-2008 α. 1.5 Millions of Rupees Difference Admin Amount

Fig. 7: Distribution of Errors in Annual Remittances 2008

