Detecting Problems in Survey Data using Benford's Law*

George Judge
University of California at Berkeley†

Laura Schechter
University of Wisconsin at Madison‡

September 7, 2006

Abstract

It is 15:00 on Friday in Nairobi. Do you know where your enumerators are??

Good quality data is paramount for applied economic research. If the data underlying economic analyses are distorted, the corresponding conclusions and inferences may be false or incomplete. We demonstrate how Benford's Law can be used to test for data abnormalities including enumerator and respondent distortions. We conduct a detailed analysis of one dataset from rural Paraguay, as well as a broader comparison of seven datasets commonly used by development economists. Using the Paraguay data, we find that Benford's Law is a useful tool with which to evaluate and test the integrity of survey data. Using seven datasets, we then compare those collected under the supervision of academic researchers and those collected by government statistical offices and present evidence regarding relative data quality.

Keywords: Benford's law, first-digit phenomenon, relative frequencies, expected frequencies, data errors, survey quality.
AMS Classification: Primary 62E20.
JEL classification: C10, C24.

*The order of the authors' names has only alphabetical significance. We have benefited from helpful comments from Jennifer Alix-Garcia, Wendy Cho, Joanne Lee, Ted Miguel, and Elisabeth Sadoulet. For a convenient utility to test your own data against Benford's law go to http://www.checkyourdata.com. John Morrow provided excellent research assistance and created the online Benford utility.

†Professor in the Graduate School, Berkeley CA. e-mail: `judge@are.berkeley.edu`.
‡Agricultural and Applied Economics, Madison WI. e-mail: `lschechter@wisc.edu`.

1 Introduction

In developing countries, much of the social and economic data are collected by surveys. Horror stories are common in which, halfway through the data analysis process somebody discovers that one (or more) enumerator is answering the survey himself rather than actually interviewing households. Also prevalent are stories in which, after spending a large sum of money to buy a data set, a researcher realizes that the information of interest to him seems inaccurate or contains large measurement errors. Since information contained in survey data often plays a key role in policy decisions, it is important to have a basis for identifying the quality of the reported data.

In data obtained from economic surveys, questions usually arise pertaining to: i) the quality of the enumerators (what if an enumerator completes the survey questionnaire while enjoying coffee at Starbucks?) and ii) the quality of the responses from those interviewed (what if the questionnaire is poorly designed and elicits answers from respondents that are inconsistent with the objectives of the question?). If either the error of omission or commission occurs, it would be useful to identify them early in the research process. Therefore, a basis upon which one could recognize survey data irregularities, manipulated outcomes, and abnormal digit and number occurrences, would be a valuable tool for researchers designing and using survey data. In this paper, we demonstrate the use of Benford's first significant digit (FSD) law as one such possibility.

In discussing the many sources of errors in survey data, Deming (2006) notes that "respondent bias and questionnaire construction are outstanding problems toward which statistical research must be directed."[1] Deaton (2005) discusses sources of error in both survey and national-accounts data and elaborates reasons why the two may not be in accord. He concludes that survey data rather than national-accounts data are necessary to analyze whether economic growth reaches the poor. At the same time, he calls for a "deeper study of the effects of nonsampling errors." Failure to recognize these data issues can lead to faulty inferences in the analysis stage.

Philipson & Malani (1999) posit that economists tend to pay more attention to the consumption of data rather than the production of data. This is evidenced by the large literature on how to deal with measurement error

[1] Attrition is another source of errors in survey data. This is discussed in detail by Thomas et al. (2001)

and the relatively small literature on how to prevent it. Philipson & Malani (1999) show how a system of random monitoring with monetary payoffs to enumerators giving errorless answers may be used to improve data collection. They work with a situation in which the data (given by doctors) can be directly verified (by the hospital). When data are not directly verifiable Benford's Law may be an alternative form of verification.

To illustrate the use of Benford's FSD law to evaluate enumerator and respondent performance, we carry out a detailed analysis on surveys from rural Paraguayan households. Using this data we find that some enumerators and questions yield higher quality responses than others. We also compare data on crops which are more important for a household's income (with importance defined in multiple ways) with data on less important crops. We find that the former are fairly well in accord with Benford's law, while the latter are much more inconsistent with Benford's law. This suggests that Benford's FSD law holds when crop quantities are more salient and so farmers are able to provide their answers with more accuracy.

In addition, we conduct a less detailed analysis on seven household surveys across the globe which have been used extensively by research economists. Using this data we also compare the quality of data collected by government statistical bureaus with that collected by academic economists.

The paper is organized as follows: Section 2 describes Benford's law; Section 3 discusses the Paraguayan data used in this paper and applies Benford's law to both the 2002 and 1999 rounds of data; Section 4 expands the analysis to compare data sets from around the globe; Section 5 discusses the implications of our results for theory and practice; and Section 6 summarizes the implications of the first-digit phenomenon in survey work.

2 Benford's Law

Benford's law characterizes the distribution of first significant digits (FSD) observed in large sets of data.[2] In 1881 Simon Newcomb observed that numbers with a first digit of 1 were observed more often than those starting with 2, 3, and so on. Newcomb was able to calculate the probability of a number having a particular nonzero first digit and published this in an article in *The*

[2] The first significant digit is the first non-zero digit when reading a number from left to right.

American Journal of Mathematics. Benford, unaware of Newcomb's article, made the same observation and published an article in *The Proceedings of the American Philosophical Society* in 1938. This FSD phenomenon was christened as Benford's Law.

Newcomb observed the probability of a number having a particular non-zero first digit as roughly

$$\Pr(\text{First digit is } d) = \log_{10}(1 + \frac{1}{d}) \tag{1}$$

where $d = 1, 2, \ldots, 9$. This formula suggests that the quantities expressed in base 10 will be uniformly distributed on a logarithmic scale. Using Newcomb's formula, the probability that the first digit of a number is 1 is about thirty percent ($P(1) = \log_{10}(1 + \frac{1}{1}) = \log_{10}(2) \approx .30$) while the probability the first digit is 9 is 4.6 percent as shown in Table 1. Thus, first significant digits typically follow a logarithmic and non-uniform distribution now identified as Benford's law.

Hill (1998) extended Newcomb's equation (1) to include digit combinations.

$$P(d_1, d_2, \ldots, d_k) = \log_{10}(1 + (d_1 d_2 \ldots d_k)^{-1})$$

For example, the probability that a number starts with the three digits 314 is $\log_{10}(1 + (314)^{-1})$. Hill & Schürger (2005) further extended this analysis to properties of the 'least significant digit' (i.e., the final non-zero digit) which can also be used to detect fraud. An interesting principle regarding Benford's law is that it seems to be scale and base invariant (Raimi 1976). Thus, multiplying data in accord with Benford's law by any positive scalar should lead to data which is also in accord with Benford's law.

Like the surprising golden ratio (Livio 2002), theories abound as to the basis of the first-digit phenomenon. Benford believed this so-called 'law' to be a general law of nature related to the logarithmic character of natural phenomena. Given Benford's conjecture, the following questions arise: i) how general is this phenomenon? and ii) what is the basis for a non-uniform FSD distribution?

As for the first question, although there are exceptions, the monotonically decreasing FSD distribution has been demonstrated to hold in data sets that include the populations of towns, budgetary data of corporations, and the half-lives of radioactive atoms. In response to the second question, there have been many attempts over the years to explain the logarithmic formula and

Table 1: Expected Frequencies Based on Benford's Law

Digit	1st place	2nd place	3rd place	4th place
0		0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.19882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982

to provide a theoretical basis for the observed phenomenon. Given that only some data sets follow the law, Hill (1995) provided a statistical derivation of the law in the form of a Central Limit Theorem for significant digits: "If distributions are selected at random and random samples are taken from each of these distributions, the significant first digits of the combined sample will converge to the logarithmic (Benford) distribution."[3] For overviews of the history and a sampling of the empirical and theoretical results, the reader is directed to Raimi (1976), Diaconis (1977), Schatte (1988), Hill (1995), Rodriguez (2004), Hill & Schürger (2005) and Berger & Hill (2006).

Since there appear to be many occurrences of Benford's law in real-life data, our objective is to exhibit data sets that may be expected to obey Benford's law and evaluate the behavioral basis of departures. Some of the others who have used Benford's law to check the validity of purported scientific data in the social sciences include Varian (1972), Carslaw (1988), Nigrini (1996, 1999), Durtschi et al. (2004), Geyer & Williamson (2004), de Marchi & Hamilton (2006) and Giles (2006).

3 Paraguayan Survey Data

To illustrate the usefulness of Benford's law, we use survey data from rural Paraguay. We find that some survey questions elicit more errors than others.

[3]This result does require some assumptions and can be thought of as a heuristic.

Table 2: Benford's Law and Aggregate Data

Variable	Obs.	1	2	3	4	5	6	7	8	9
Benford		30.10	17.61	12.49	9.69	7.92	6.69	5.80	5.12	4.58
Income	222	28.38	16.67	7.66	12.61	10.36	9.01	6.31	4.05	4.95
Ag. Income	218	22.94	16.06	10.55	14.22	7.80	7.80	7.34	6.88	6.42
Theft	112	33.04	15.18	10.71	7.14	12.50	3.57	8.04	5.36	4.46
Gifts Given	176	27.84	18.18	18.75	8.52	10.80	3.98	5.11	2.84	3.98

For example, questions for which farmers know the exact answer generally conform to Benford's law while questions for which farmers may be unsure of the exact answer do not seem to conform. For these questions respondents are much more likely to choose numbers starting with 5 than Benford's law would suggest. Finally, questions about which farmers may have an incentive to hide the truth, such as their donations to church, do not conform with Benford's law. We also show that data collected by some enumerators are in accord with Benford's law while data collected by others are not.

In 1991, the Land Tenure Center at the University of Wisconsin in Madison and the Centro Paraguayo de Estudios Sociológicos in Asunción worked together in the design and implementation of a survey of 300 rural Paraguayan households in 16 villages in three departments (comparable to states). The sampling design resulted in a random sample, stratified by land-holdings. The original survey was followed by subsequent rounds of data collection in 1994, 1999, and, most recently, in 2002. Summary statistics about both the households and the respondents can be found in Table A-1 in Appendix A.

3.1 Benford's Law Applied to 2002 Paraguay Data

3.1.1 Aggregated Data

We now examine via goodness-of-fit tests whether the data collected in Paraguay conforms to Benford's law. Initially we analyze aggregated data not directly recorded by enumerators, but from manipulations of variables recorded by enumerators. Figure 1 shows how the data on annual income compares with Benford's law. The data in Table 2 indicates how aggregated data on total income, agricultural income, theft experienced, and gifts given compare with Benford's Law. From a review of the table, it appears the data conforms to Benford's Law.

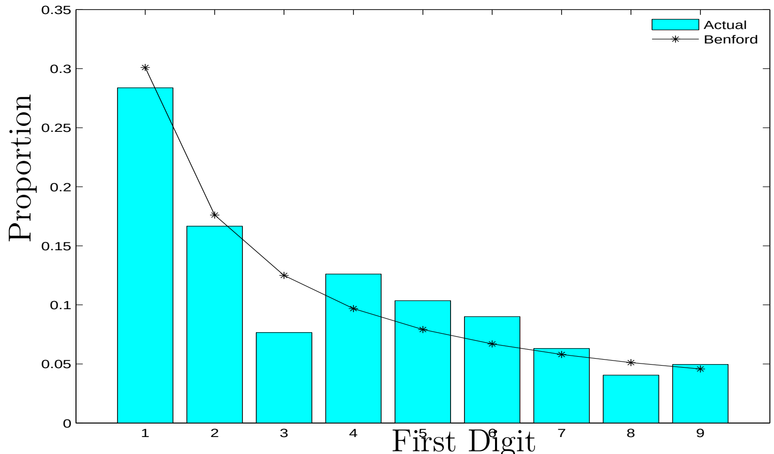

Figure 1: Annual Total Income

In addition to merely visually reviewing the data, it is possible to use a statistic to measure the extent to which data conforms with Benford's Law. How to best evaluate the empirical data is no small problem. The χ^2 goodness-of-fit test is the test most commonly used when comparing actual data with discrete expected outcomes such as Benford's Law. The χ^2 statistic is calculated as $\chi^2 = \sum_{i=1}^{9} \frac{(E_i - B_i)^2}{B_i}$ with 8 degrees of freedom where E_i is the observed frequency in each bin in the empirical data and B_i is the frequency expected by Benford. Unfortunately, χ^2 tests have enormous power for large samples and so even quite small deviations from Benford's law will be statistically significant. Thus, the χ^2 test will be more likely to reject that the data is distributed according to Benford's law for larger samples than for smaller samples. If E_i and B_i are the frequencies in each bin while e_i and b_i are the proportions, then $\chi^2 = \sum_{i=1}^{9} N \frac{(e_i - b_i)^2}{b_i}$. One can clearly see how the statistic is directly related to sample size.

Giles (2006) suggested using Kuiper's modified Kolmogorov-Smirnov goodness-of-fit test (V_N) instead because it may be less sensitive to sample size and also recognizes the circularity of the data.[4] Kuiper's V_N statistic is calculated as $V_N = \max_x[F_e(x) - F_b(x)] + \max_x[F_b(x) - F_e(x)]$ where $F_e(x)$ is the empirical CDF of the FSD distribution and $F_b(x)$ is Benford's CDF. Critical values for a modified Kuiper test ($V_N^* = V_N[N^{1/2} + 0.155 + 0.24N^{-1/2}]$) have been given by Stephens (1970). However, both the original and modified

[4]The fact that FSD of 9 and 1 are actually close to one another.

Table 3: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Aggregate Data

Variable	r	m	χ^2	V_N^*
Income	0.95	0.048	10.55	1.23
Ag. Income	0.96	0.072	13.54	1.59*
Theft	0.95	0.046	7.31	0.72
Gifts Given	0.94	0.063	11.94	1.15

* indicates 95% and ** indicates 99% significantly different from Benford.

Kuiper tests were designed for use with continuous distributions, so the critical values given by Stephens (1970) are inaccurate for the discrete Benford distribution. Instead, Monte Carlo exercises suggest a 5% critical value of 1.34 for the Benford distribution as opposed to 1.75 in the continuous case. The 10%, 5%, and 1% critical values for χ^2 are, respectively, 13.36, 15.51, and 20.09, and for V_N^* they are approximately 1.21, 1.34, and 1.61.

We have conducted our own power tests and find that the Kuiper test also exhibits increasing power with an increase in the number of observations. If the data is distributed as the absolute value of a normal with a mean of 0 and standard deviation of 1, then the power of the 99% significance test (the probability of rejecting the null hypothesis that the data is distributed according to Benford) with 300 observations is only 82%. With 500 observations the power increases to over 97%. If, instead of being normally distributed, the data is distributed as is the crop production data in 2002 in Paraguay, then the power of the 99% significance test with 300 observations decreases to 22%. With 1000 observations the power increases to over 97%. This evidence suggests that low power will not be a problem for data sets with over 1000 observations.

It becomes more problematic to compare the quality of data sets of different sizes when tests find the data are not in accord with Benford's law. Both the χ^2 and V_N^* statistics tend to increase with both the number of observations and the distance between the data and Benford.[5] For this reason we continue to include the correlation between the data and Benford as another measure of compatibility with Benford's law. Additionally, as suggested by Leemis et al. (2000), we calculate a statistic called $m = \max_{i=1,2,\ldots,9}\{|b_i - e_i|\}$. This

[5]Nigrini & Miller (2006) have similar problems testing conformance with Benford's law on large data sets of hydrological phenomena.

Table 4: Benford's Law and Production Quantities

Variable	Obs.	1	2	3	4	5	6	7	8	9
Benford		30.10	17.61	12.49	9.69	7.92	6.69	5.80	5.12	4.58
All Products	1632	31.50	19.30	12.81	7.90	12.93	5.82	4.11	3.62	2.02
Enumerator 1	516	30.23	19.19	13.37	8.72	10.66	8.91	2.71	4.46	1.74
Enumerator 2	556	32.73	19.06	12.59	7.73	12.77	6.29	3.42	3.42	1.98
Enumerator 3	560	31.43	19.64	12.50	7.32	15.18	2.50	6.07	3.04	2.32
P.I. sat in	582	33.33	16.49	13.06	7.04	13.92	4.64	4.47	4.64	2.41
P.I. didn't sit in	1050	30.48	20.86	12.67	8.38	12.38	6.48	3.90	3.05	1.81

statistic is also less sensitive to sample size.[6] Morrow (2006) shows that there are general properties under which we should expect Benford's law (and scale invariance) to hold, however he also shows that the suitability of tests found in the literature is dependent on underlying distributional assumptions.

All of these statistics are given in Table 3. Note that the relative FSD frequencies found in the data are highly correlated with Benford's distribution and we cannot, on the basis of the χ^2 and V_N^* goodness-of-fit tests, reject the hypothesis that the aggregated Paraguayan survey data follows the Benford distribution at the 95% significance level.

3.1.2 Enumerated Data

In addition to testing the aggregated data, we also test the data for which the enumerator directly recorded the answer as stated by the farmer. The farmer was asked how much of each crop was harvested by his household in the past year. The quantities produced of all possible crops were compiled as shown in Appendix A in Figure A-1.[7] In Table 4 we find that quantities with an FSD of 5 are much more common than suggested by Benford's Law. Accordingly, the goodness-of-fit tests reported in Table 5 suggest a rejection of the hypothesis that the data on quantities produced was generated by

[6] Note that higher values of the Kuiper V_N^* test, the χ^2 test, and the m statistic imply the empirical data is less similar to Benford's distribution, while higher values of the correlation imply the data is more similar to Benford's distribution.

[7] If a household produces 150 kilos of corn and 420 kilos of cassava that will count as FSDs of 1 and 4, not 5.

Table 5: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Production Quantities

Variable	r	m	χ^2	V_N^*
All Products	0.97	0.050	101.34**	2.69**
Enumerator 1	0.97	0.031	28.20**	1.50*
Enumerator 2	0.98	0.049	37.58**	1.68**
Enumerator 3	0.94	0.073	67.92**	1.97**
P.I. sat in	0.96	0.060	44.94**	1.47*
P.I. didn't sit in	0.97	0.045	67.54**	2.26**

* indicates 95% and ** indicates 99% significantly different from Benford.

Benford's distribution.[8]

We might worry that one of the three enumerators was not completing surveys properly, thus explaining why the quantity data does not conform with Benford's law. Moving in this direction, we divide the data on quantities produced by each of the three enumerators who worked on the survey. In addition, the Principal Investigator (P.I.) sat in on interviews with a different enumerator every day, alternating between the three. There was no specific type of household the P.I. tended to visit more often. We analyze whether or not the distribution of FSD changes based on the identity of the enumerator and the presence of the P.I. during the interview. Results suggest that bias remains, although we cannot reject that the results given by the first enumerator and from surveys at which the P.I. was present are in accord with Benford's law. (This is somewhat reassuring for the P.I.'s ego.)

As these data are not falsified, the relative frequency departures are probably due to the fact that the farmers were not always sure exactly how much of a crop they had harvested and so tended to choose 'nice' round numbers, e.g. a farmer is more likely to claim a harvest of 500 kilos of corn than 422 kilos. From the results in Table 5, we can reject at the 99% level the hypothesis that the production data were generated in accord with Benford's law. One could argue that the overabundance of observations starting with 5 is not due to guesstimation but rather to the fact that farmers' plots are similar sizes and thus generate crop production of similar amounts. If this were the

[8] de Marchi & Hamilton (2006) use Benford's law to test for tampering in self-reported toxics emissions by plants and likewise find that there are a disproportionate number of quantities beginning with 5.

Table 6: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Other Items

Variable	r	m	χ^2	V_N^*
Animals Owned	0.99	0.044	18.52*	1.14
Has. Land	0.98	0.034	8.49	0.72
Church Donations	0.92	0.083	38.93**	1.64**

* indicates 95% and ** indicates 99% significantly different from Benford.

case, we would expect that there should also be relatively many observations beginning with 4 and 6 as well. However, Figure A-1 in Appendix A shows that this is not the case - the abundance of 5's is at the expense of 4's and 6's.

This illustrates how Benford's law can be used to distinguish problems in survey data arising because of an enumerator from those arising because of an ill-phrased question. For researchers who design surveys and collect data, this situation can be identified early in the survey collection process to help avoid major difficulties in subsequent econometric analysis.

3.1.3 Other Variables

In Appendix A, Table A-2, we perform the same exercise on a few variables also obtained directly from the respondents which seem likely to follow Benford's law. From the results in Table 6, we cannot reject that the data on the number of animals owned, and hectares of land owned or used come from Benford's distribution using the V_N^* test. The animals-owned variables include separately the quantities owned of each animal species. If households are able to report the assets they own more accurately than the quantities of crops harvested, measures of wealth for rural households in developing countries may be more accurate than measures of income. For research in which either wealth or income can be used, this may suggest a preference for the former.

We reject the hypothesis that donations to church are in accord with Benford's Law. This may be because respondents are not sure how much they donated to church or because respondents are reluctant to answer honestly. From an economic standpoint it is important to know the main cause of the departures from Benford's distribution.

In Figures 1, A-1, and A-2 we compare histograms of the actual distri-

Table 7: χ^2 Tests and Kuiper V_N^* Tests between Benford's Law and Animals Owned and Crops Produced

	# Surveys	# Obs	Obs/Survey	χ^2	V_N^*
Animals Owned (1999)	298	827	2.8	56.21^{**}	3.09^{**}
Animals Owned (2002)	223	676	3.0	18.52^{*}	1.14
Crops Produced (1999)	298	1412	4.7	30.36^{**}	1.21
Crops Produced (2002)	223	1632	7.3	101.34^{**}	2.69^{**}
P.I. sat in (2002)	71	582	8.2	44.94^{**}	1.47^{*}
P.I. didn't sit in (2002)	152	1050	6.9	67.54^{**}	2.26^{**}

* indicates 95% and ** indicates 99% significantly different from Benford.

butions for total income, production quantities, and the number of animals owned. They all appear quite similar to Benford's distribution. The distribution for production quantities does not look as dissimilar to Benford's distribution as the χ^2 goodness-of-fit test suggested because the histograms disguise the fact that there were over 1000 observations on production quantities.

3.2 Benford's Law Applied to 1999 Paraguay Data

The Paraguay data comes from a panel data survey collected in 1991, 1994, 1999, and 2002. Although neither of the authors worked on data collection in the previous rounds, we do have access to those data sets. In 2002 only three enumerators worked on the survey, while in 1999 ten enumerators worked on the survey. Having so many enumerators with only one supervisor might increase the probability of enumerators making up answers because of less supervision during surveys. We test if this is the case.

We have seen that the number of animals owned recorded in 2002 fit Benford's distribution well. We can test whether the same is true of the 1999 data. Table A-3 gives a detailed analysis of animals-owned data by enumerator in both rounds of the surveys. Table 7 contains a more succinct analysis. Using the 2002 data we cannot reject that the distribution of the responses from any of the enumerators differed from the Benford distribution, but the data from 1999 is more ambiguous - some enumerators seem to do a good job using the Benford benchmark while others' data are not consistent with Benford's distribution.

We have learned from analyzing the 2002 data on quantities of crops produced that it is subject to much respondent error. As a basis of comparison, we analyzed the 1999 data. Table A-4 provides details regarding how the enumerators in 1999 and 2002 fared in terms of the answers recorded for quantities of crops produced, and Table 7 contains a more succinct analysis. Note that the production-quantities data in 2002 looks more suspect than the data in 1999. This result is surprising at first, given the previous result that the animals-owned data was better in 2002, but the contradiction can be explained.

Enumerators in 1999 recorded fewer crop-production quantities per household. In 1999 households were recorded as producing an average of 4.7 different crops which rose to 7.3 in 2002. This could be due to an increase in diversification in the three years from 1999 to 2002. The likelier case is that, in 2002, enumerators were encouraged to be quite comprehensive and collect data on all crops produced, not just the most important ones.[9] Respondents may not be sure about the exact quantity produced of crops which are less important to their livelihood (Groves 1989). This emphasizes the need for caution in using Benford's Law. Although quantities of crops produced may be reported less accurately for less important products, ignoring them altogether will not increase the accuracy of measures of total income. This also leads to a bias in the sense that the income of more diversified farmers will contain more error than the income of relatively less diversified farmers.

An alternative explanation is that, by asking farmers for a more comprehensive list of crops planted, they lost patience with us and stopped answering the questions with as much thought. This could also lead to the lower quality crop production data in 2002. We test this hypothesis by comparing data on 'important' crops to that on 'non-important' crops. This is shown in Table 8 using three definitions of importance. First, we compare total quantities harvested of crops, which were sold by the household, with those grown for home consumption. Next, we compare crops whose harvests were worth more to the household than 500,000 Guaranies in 2002 (342,500 Guaranies or more in 1999 accounting for inflation, a bit less than $100) with those which were worth less. Lastly, we compare the four most valuable crops (in terms of the value of total output) for each household with any additional, less valuable,

[9] This is supported by the fact that, in 2002, the average number of different crops produced as recorded by the enumerator when the P.I. sat in on the survey was 8.2 versus 6.9 when the P.I. did not sit in on the survey.)

Table 8: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests for More and Less 'Important' Crops

	# Obs	r	m	χ^2	V_N^*
Crops Sold in 1999	378	0.99	0.027	5.58	0.72
Crops Not Sold in 1999	1034	0.98	0.036	40.12**	1.51*
Crops Sold in 2002	384	0.97	0.038	12.75	0.82
Crops Not Sold in 2002	1248	0.97	0.054	100.82**	2.87**
Crops Value$>342,500$ in 1999	889	1.00	0.018	9.81	0.58
Crops Value$\leq 342,500$ in 1999	523	0.95	0.066	53.81**	1.57*
Crops Value$>500,000$ in 2002	667	0.99	0.027	10.54	1.17
Crops Value$\leq 500,000$ in 2002	965	0.94	0.085	164.08**	3.34**
Hh's top 4 crops in 1999	1026	0.99	0.019	13.73	0.77
Hh's other crops in 1999	386	0.98	0.045	28.53**	1.47*
Hh's top 4 crops in 2002	828	0.98	0.032	27.25**	1.04
Hh's other crops in 2002	804	0.96	0.069	95.08**	2.78**

* indicates 95% and ** indicates 99% significantly different from Benford.

crops.[10]

Looking at Table 8 we see that 'important' crops, no matter how defined, are always more in accord with Benford's law than those defined as 'less important'. These results are quite striking given the evidence that these tests usually become more conservative as sample size increases. There are usually more observations for 'non-important' crops than important ones, no matter how defined, in 2002 than in 1999. On the other hand, there were usually more 'important' crops enumerated in 1999 because the 1999 sample includes 298 households whereas the 2002 sample only includes 223. This suggests that the enumerators in 2002 included more of households' secondary crops. As such, the inclusion of these less important crops, rather than respondent fatigue, seems to be why the 2002 crop data appears to be of lower quality than the 1999 crop data.[11]

[10] If a household planted fewer than four crop varieties we included all crops in the 'valuable' category.

[11] More important crops, about which there is probably less guessing, conform more closely to Benford's law. This may be used as evidence to help convince those skeptical of Benford's ability to describe the distribution of FSD in naturally occurring data.

4 Comparing High-Profile Data Sets

In this section we analyze the quality of seven data sets which have been used frequently in academic publications. See Appendix B for a list of the many papers published using these data sets. In the process, we compare the quality of data collected by government statistical bureaus with that collected by development economists. In recent years, it has become more popular for researchers to supervise their own data collection, but nothing is known about the relative quality of these homegrown data sets. *A priori* one could argue why either should be of higher quality.

We look at two data sets collected under the supervision of academic economists and five data sets collected under the supervision of government agencies. We have no prior beliefs about which type of data set should be of higher quality. One could argue that data collected under the supervision of academics should be of higher quality because the supervisor has a greater personal interest in the quality of the data and will thus monitor the enumerators more closely. These data sets also tend to be smaller, which may increase the potential for monitoring. On the other hand, data collected under the supervision of government agencies and the World Bank may be of higher quality because they are professional survey administrators and may have more experience and knowledge about surveying techniques. In addition, those in charge are often from the country in which they are gathering data and so may be less likely to have misunderstandings in the field.

In this study, the data sets collected by academic researchers seem to be more free of distortions than those collected by the government. As these data sets have been used in a multitude of academic papers and have been used to make policy prescriptions, the importance of data issues should be kept in mind.

The first of the two data sets collected by academic economists and discussed here is the Paraguayan data discussed thus far. The 2002 round of data was collected under the supervision of one of the authors when she was a graduate student at UC Berkeley and includes 223 households. The 1999 round of data was collected under the supervision of a professor who was then a graduate student at UW Madison and includes 298 households. The second data set we will analyze was collected in Ghana from 1996-1998 under the supervision of Chris Udry (a professor at Yale) and Markus Goldstein (then a graduate student at UC Berkeley). The data set includes information on 294 households.

The data collected by government statistical bureaus with the help of international organizations contain many more observations. The five data sets we examine are:

1. The Matlab Health and Socioeconomic Survey (MHSS) was collected in 1996 as a collaborative effort by RAND, multiple universities in the United States, and research centers in Bangladesh. There is data on over 4,500 rural Bangladeshi households.

2. The Progresa data from Mexico consists of panel data for 24,000 rural Mexican households collected every 6 months beginning in November of 1997. This data was collected by Progresa, which is part of the Mexican government, with consultations from the International Food Policy Research Institute (IFPRI).

3. The IFPRI Pakistan data includes 14 rounds of panel data covering rural households and villages spanning 1986-91. We have used all rounds except the thirteenth. The survey was jointly produced by IFPRI, the Government of Pakistan, and the U.S. Agency for International Development (USAID). Large fluctuations in agricultural production observations across rounds are in part due to the particular season (monsoon or winter) in which that round was conducted.

4. The Perú Living Standards Measurement Survey (LSMS) (called PLSS or ENNIV) contains information on both rural and urban households. Although a few of the urban households own animals or grow crops, we have excluded these households to maintain comparability with the other data sets which only interviewed rural households. The survey has data on 2,349 rural households in 1985, 594 rural households in 1991, and 1,336 rural households in 1994. The 1985 data was collected by Statistical Institute of Perú ("Instituto Nacional de Estadística e Informática del Perú (INEI)"), with technical and financial support from the World Bank and the Central Reserve Bank of Peru. The 1991 and 1994 data were collected by the Peruvian research enterprise Cuánto S.A. with technical and financial assistance from the World Bank (with additional assistance from the Interamerican Development Bank in 1994).

5. The Vietnam Living Standards Survey (VLSS) contains information on 4,800 households in 1992 and 6,000 households in 1998. This data was

collected by the General Statistical Office of Vietnam with help from the World Bank.

We have chosen to look at two variables which are likely to be in accord with Benford's law and are comparable across surveys, the quantities of crops harvested and the number of animals owned. We expect these data to be in accord with Benford's law for two reasons, the first being that they concern, arguably, naturally occurring phenomena. Although farmers plant seeds, nature conditions the output, and although farmers sell and buy animals, they reproduce at their own pace. Secondly, both variables are combinations of distributions. Observations of crops produced are the combination of distributions for all the crops that households could produce while the animals-owned variable combines the distributions for all the types of animals households could own.

Table 9 shows the results for crop production quantities in the seven different surveys. We see that the χ^2 statistic rises with sample size, a disadvantage we noted previously. On the other hand, the V_N^* test is less affected by sample size. We find that the data sets collected by academic researchers contain data that is much more in accord with Benford's Law than the government-collected data sets, even when making comparisons with the smaller government data sets (all rounds of the Pakistan data and the data from Peru in 1991).[12]

Within the data collected by the government and international organizations, the Progresa data from Mexico and the IFPRI data from Pakistan seem to be the least in accord with Benford's law. When looking at the correlation and the m statistic, measures which are less affected by sample size, the Mexico and Pakistan data are the only ones with correlations below .90 and m statistics above 0.10. Although the χ^2 and V_N^* test statistics are related to sample size, we can compare the data from Mexico with that from Vietnam, Peru in 1985, and Bangladesh. The test statistics for the data from Mexico are much higher than those in the other three data sets. The test

[12]We asked Markus Goldstein, one of the supervisors of data collection in Ghana, why he thought the data they collected was of such high quality. Markus replied that, as do many other organizations, they paid enumerators high wages so they would be afraid of losing their jobs. More unusually, they made a point of hiring almost all high school graduates, with only one college graduate. Lastly, Chris and Markus tried to be in the field a few days a week, sitting in on interviews for an average of one and a half days a week.

statistics for the Pakistan data are more similar to those of the other smaller data sets.

Table 10 shows each survey's results for the number of each type of animal owned. Comparing the quality of data on crop production (a component of income) with that on animals owned (a component of wealth), neither seems to be clearly better in all of the data sets.

Table 9: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Crop Quantities Produced

	# Surveys	# Obs	r	m	χ^2	V_N^*
Ghana (96-98)	294	799	0.99	0.016	8.45	0.85
Paraguay (1999)	298	1,412	0.99	0.026	30.36**	1.21
Paraguay (2002)	223	1,632	0.97	0.050	101.34**	2.69**
Bangladesh (1996)	4,522	7,357	0.93	0.065	440.92**	8.27**
Mexico (Oct 98)	24,067	10,028	0.84	0.154	3959.21**	15.45**
Mexico (May 99)	22,328	9,659	0.88	0.134	3145.46**	13.14**
Pakistan (July 86)	927	874	0.97	0.040	31.62**	1.84**
Pakistan (Oct 86)	909	181	0.86	0.100	23.37**	1.40*
Pakistan (Jan 87)	882	563	0.99	0.047	20.71**	1.13
Pakistan (Mar 87)	854	46	0.94	0.051	4.79	0.73
Pakistan (Apr 87)	845	191	0.96	0.041	15.90*	1.13
Pakistan (July 87)	831	658	0.97	0.033	43.75**	0.95
Pakistan (Jan 88)	813	728	0.96	0.046	41.45**	1.41*
Pakistan (Mar 88)	809	69	0.92	0.048	8.11	1.16
Pakistan (Aug 88)	804	559	0.99	0.022	15.44	0.77
Pakistan (Jan 89)	802	708	0.99	0.034	26.23**	1.27
Pakistan (Mar 89)	766	200	0.78	0.100	29.67**	2.11**
Pakistan (Aug 89)	759	691	0.98	0.025	22.48**	1.06
Pakistan (Oct 91)	726	1,095	0.99	0.022	18.72*	1.16
Peru (1985)	2,349	8,050	0.99	0.026	287.76**	4.23**
Peru (1991)	594	2,495	0.98	0.047	228.02**	5.47**
Peru (1994)	1,336	4,101	0.98	0.038	246.39**	4.63**
Vietnam (1992)	4,800	20,936	0.99	0.034	762.36**	7.80**
Vietnam (1998)	6,002	24,420	0.98	0.035	1092.93**	9.64**

* indicates 95% and ** indicates 99% significantly different from Benford.

Table 10: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned

	# Surveys	# Obs	r	m	χ^2	V_N^*
Ghana (Nov 96)	294	332	0.99	0.023	7.79	0.79
Ghana (Dec 97)	294	335	0.99	0.016	2.76	0.60
Ghana (Aug 98)	294	306	0.99	0.017	3.83	0.75
Paraguay (1999)	298	827	0.93	0.061	56.21**	3.09**
Paraguay (2002)	223	676	0.99	0.044	18.52*	1.14
Bangladesh (1996)	4,522	6,807	0.99	0.057	333.56**	7.25**
Mexico (Nov 97)	24,077	48,042	0.99	0.021	1142.88**	6.64**
Mexico (Oct 98)	24,067	37,130	1.00	0.098	4052.83**	27.23**
Mexico (May 99)	22,328	34,539	1.00	0.108	4065.22**	28.15**
Mexico (Nov 99)	23,266	38,817	0.99	0.114	4750.71**	30.85**
Mexico (May 00)	22,627	37,580	1.00	0.115	4703.09**	29.88**
Pakistan (July 86)	927	1,562	0.96	0.119	358.08**	8.07**
Pakistan (Oct 86)	909	1,981	0.94	0.120	386.43**	7.36**
Pakistan (Jan 87)	882	2,010	0.77	0.128	395.04**	8.82**
Pakistan (July 87)	831	1,961	0.84	0.116	326.96**	7.95**
Pakistan (Mar 88)	809	1,786	0.77	0.171	456.86**	8.58**
Pakistan (Mar 89)	766	1,838	0.81	0.154	402.10**	8.28**
Pakistan (Oct 91)	726	559	0.71	0.135	89.93**	3.83**
Peru (1985)	2,349	8,007	0.98	0.072	696.76**	8.73**
Peru (1991)	594	2,369	0.97	0.069	220.62**	4.82**
Peru (1994)	1,336	3,392	0.99	0.048	126.40**	3.72**
Vietnam (1992)	4,800	8,005	0.99	0.102	1372.73**	16.79**
Vietnam (1998)	6,002	8,351	0.97	0.092	1154.57**	13.95**

* indicates 95% and ** indicates 99% significantly different from Benford.

The results for the animals-owned data also indicate that the data collected by academics seem to be more in accord with Benford's law than that collected by government agencies. Again, the correlations are lowest for the Pakistan data, and the m statistics are highest for the data from Pakistan and Mexico. We still find that, in terms of tests, the data from Mexico compares poorly with that from Vietnam and Peru while the data from Pakistan compares poorly with that from Bangladesh and Peru.

Note that the November 1997 round of data from Mexico appears to be of much higher quality than the other rounds. (Data on crop production was not collected in this round, so we cannot make the similar comparison in the previous Table 9.) This is interesting because the first round of the Mexican Progresa data has a different survey name (Encaseh for determining household eligibility vs Encel for program evaluation). It is not clear why the difference between the two surveys should be so large.

4.1 Male vs Female Respondents

Four of the surveys identify which household member responded to the questionnaire. If women are in charge of livestock while men are in charge of crop production, one might think that women can answer questions about livestock more accurately than men, while men can answer questions about crop production more accurately than women. In Tables 11 and 12 we test this idea. The Bangladesh, Peru, and Vietnam data indicate the respondent's identity. For the 2002 Paraguay data, some surveys were conducted with multiple family members so there are surveys answered by males, females, and both.

The test results are always much higher (worse) for males than females, but that may be because of the higher sample size of male respondents. The correlations and m statistics, which do not depend on sample size, are quite similar for males and females. These results suggest that there is not a large difference in the overall quality of information given by male versus female respondents. Additionally, males and females do not seem to perform differently in answering questions related to crops versus livestock. We might interpret these results as showing that survey supervisors can get reasonable quality data from both male and female household members, although this ignores the fact that households endogenously chose which household member would answer the survey questions.

Table 11: Gender: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Crop Quantities Produced

	Sex	# Surveys	# Obs	r	m	χ^2	V_N^*
Paraguay (2002)	Male	138	997	0.98	0.050	65.76**	2.24**
Paraguay (2002)	Female	43	294	0.96	0.057	22.64**	1.12
Paraguay (2002)	Male and Female	42	341	0.97	0.044	20.44**	1.24
Bangladesh (1996)	Male	3,546	6,481	0.93	0.066	389.42**	7.98**
Bangladesh (1996)	Female	976	876	0.94	0.058	63.01**	2.49**
Peru (1985)	Male	1,632	6,750	0.99	0.025	211.44**	3.57**
Peru (1985)	Female	500	1,300	0.99	0.030	42.50**	1.66**
Peru (1994)	Male	1,026	3,546	0.98	0.048	197.86**	3.97**
Peru (1994)	Female	180	555	0.96	0.047	57.27**	2.57**
Vietnam (1992)	Male	2,648	14,629	0.99	0.035	565.90**	7.00**
Vietnam (1992)	Female	1,390	6,307	0.99	0.033	210.37**	3.63**
Vietnam (1998)	Male	2,763	15,485	0.99	0.034	696.09**	8.06**
Vietnam (1998)	Female	1,671	8,922	0.98	0.036	402.43**	5.34**

* indicates 95% and ** indicates 99% significantly different from Benford.

Table 12: Gender: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned

	Sex	# Surveys	# Obs	r	m	χ^2	V_N^*
Paraguay (2002)	Male	138	414	0.99	0.069	17.77*	1.41*
Paraguay (2002)	Female	43	126	0.96	0.056	6.60	0.64
Paraguay (2002)	Male and Female	42	136	0.95	0.065	12.63	0.87
Bangladesh (1996)	Male	3,546	5,657	0.98	0.059	281.84**	6.34**
Bangladesh (1996)	Female	976	1,150	0.99	0.056	59.64**	3.60**
Peru (1985)	Male	1,632	6,422	0.97	0.078	563.81**	8.28**
Peru (1985)	Female	500	1,585	0.99	0.048	99.77**	3.55**
Peru (1994)	Male	1,026	2,915	0.99	0.048	105.57**	3.48**
Peru (1994)	Female	180	477	0.98	0.048	28.85**	1.88**
Vietnam (1992)	Male	2,648	5,536	0.99	0.100	913.25**	13.57**
Vietnam (1992)	Female	1,390	2,469	0.99	0.107	464.11**	9.93**
Vietnam (1998)	Male	2,763	5,338	0.97	0.090	669.87**	10.94**
Vietnam (1998)	Female	1,671	3,009	0.98	0.095	443.34**	9.00**

* indicates 95% and ** indicates 99% significantly different from Benford.

4.2 Enumerators' Opinions

Some surveys ask the enumerator to judge the quality of information given by the respondent. Using Benford's Law, we can see how well enumerators identified low-quality data. The survey from Bangladesh asked enumerators to judge both the accuracy of the respondents' answers as well as the seriousness and attentiveness of the respondent. Possible answers were: excellent, good, fair, not so bad, and very bad. In Table 13 we compare those surveys which were judged to be fair, not so bad, or very bad in terms of either accuracy or attentiveness (or both) with those that were good or excellent in both categories. As there are many more 'good' observations than 'bad', the test statistics for the good data are higher. More suggestively, the correlation with Benford is also higher in the 'bad' data while the m statistic is lower. If anything, the bad data seems to be better than the good data. Although this analysis is only applied to one data set, it suggests that enumerator evaluations of the respondents' data should be taken with a grain of salt.

Table 13: Enumerator Opinion: Correlations (r), the m Statistic, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned and Crop Quantities Produced

	Quality	# Surveys	# Obs	r	m	χ^2	V_N^*
			Crops Produced				
Bangladesh (1996)	'Bad'	1,384	2,096	0.95	0.052	94.33**	3.46**
Bangladesh (1996)	'Good'	3,107	5,204	0.92	0.070	360.55**	7.53**
			Animals Owned				
Bangladesh (1996)	'Bad'	1,384	2,053	1.00	0.046	80.56**	3.84**
Bangladesh (1996)	'Good'	3,107	4,716	0.98	0.066	262.15**	6.16**

* indicates 95% and ** indicates 99% significantly different from Benford.

5 Implications for Theory and Practice

These results suggest that measurement error is a problem in many data-collection exercises. When respondents are asked for answers of which they are unsure they tend to estimate and round to 'nice' numbers. In addition, at least in this study, larger data sets collected by government statistical offices seem to be of lower than data collected by academic researchers. The Progresa data in Mexico and the IFPRI data from Pakistan are particularly inconsistent with Benford's law. These are issues which should not be ignored. Although sophisticated econometric techniques are available to deal with measurement error once it is identified, we should be much more careful and serious about both enumerator quality and designing questionnaires that elicit data with minimal respondent errors.

One might argue that the government-collected data sets still compare quite favorably with the academic-collected data sets due to their larger sample sizes. Having data sets which are an order of magnitude larger but contain more measurement error may still allow for more power than smaller data sets with less error. This would be true if the errors were pure white noise, but one might expect that this is not the case. We have shown evidence that suggests data errors increase for more-diversified farmers. If certain questions are more prone to errors than others, then we will find that surveys for households which are more active in those areas will contain more errors, which can cause serious problems in estimation and inference.

We have also shown that there are certain questions which are more or less susceptible to response errors. Questions for which people tend to be unsure of the answer or which people may have an incentive to answer dishonestly, such as donations to church, are more susceptible to errors. The desire to produce data which minimizes these irregularities can and should influence data collection methodology.

Although the exact questions which lead to departures from Benford's FSD distribution may be different in each country and situation, Benford's law provides a simple means of testing for such irregularities in data. Researchers can quickly test whether the variable that is of most interest to their research follows Benford's law or exhibits errors in the pre-testing stage of the survey. Glewwe & Dang (2005) show how having computers available for data-input and mistake-finding at the district level, so that mistakes can be found more quickly and households reinterviewed sooner, can improve data quality. These computers could easily be programmed to include a Benford's law component, able to test for the quality of different questions and different enumerators.

One should remember that these errors are most easily tested for in data directly from the respondent. For example, as we saw above, Benford's law suggests that the data on crop quantities produced is irregular but we cannot reject that total income (the sum of all quantities produced multiplied by their prices plus other sources of income) follows Benford's law. This is because we calculated total income as a linear combination of many variables. It has been suggested that combining variables may lead them to better approximate Benford's distribution even when the underlying individual variables do not follow Benford's law (Raimi 1976).

These results demonstrate why one should not consider only Benford's Law when evaluating enumerators or data sets. For example, while the data collected in 1999 on crop quantities produced are much more in accord with Benford's law than that collected in 2002, the evidence suggests that this is not because the enumerators were better in 1999. The enumerators in 1999 seem to have only collected data on the most important crops, for which there was less respondent error, while the enumerators in 2002 collected data on many more crops, but for which there was more respondent error. Hence, while Benford's law suggests that the 1999 data contains less measurement error, other evidence suggests that this is because the 1999 data includes fewer crops. This is a warning against using Benford's law in isolation when judging the quality of a data set.

6 Conclusions

We have demonstrated how Benford's Law can be used to detect data abnormalities arising both from questions that are difficult to answer and from enumerator errors. While econometricians and applied economists spend much energy correcting for measurement error in pre-existing data sets, they should also try to avoid it by detecting these problems early in the data-collection process.

There remains much room for future research on topics related to survey design and enumerator contracts. Can researchers articulate which types of questions and situations will lead to more accurate answers in general? For example, the following situations may affect error: use of interpreters, presence of non-family members during the interview, participation of more than one family member in the interview, and participation of female rather than male household members.

Furthermore, while Philipson & Malani (1999) show how random enumerator audits with prizes for accurate reporting can be used to decrease errors when direct data verification is possible, a contract has not yet been designed for data tests such as Benford's law which may be more prone to both Type I and Type II errors. These are important steps that should be taken to increase the quality of data production in addition to that of data consumption.

Finally, Scott & Fasli (2001) note that even in Benford's original paper only half of the data sets provide a reasonably close fit with Benford's law. Consequently, it seems possible that a family of data-based FSD distributions may be more compatible with observed data sets. To this end we explore the use of information-theoretic methods in another paper to develop such a family of alternative Benford-like distributions (Grendar et al. 2006).

A Appendix: Extra Tables and Figures

Table A-1: Summary Statistics for 2002 Paraguay Data

	Household Variables	
Variable	Mean	(Std. Dev.)
Theft Experienced	111,000	(336,000)
Gifts Given	306,000	(524,000)
Annual Income	28,300,000	(72,100,000)
Median Annual Income	9,046,000	
Family Size	5.6	(2.4)
Land Owned (hectares)	36.6	(95.8)
	Respondent Variables	
Variable	Mean	(Std. Dev.)
Male	79%	
Age	52.2	(14.8)
Years of Education	4.9	(2.7)
Obs	223	

The relevant exchange rate is approximately 4,800 Guaranies to the dollar.

Table A-2: Benford's Law and Other Paraguayan Items in 2002

Variable	Obs.	1	2	3	4	5	6	7	8	9
Benford		30.10	17.61	12.49	9.69	7.92	6.69	5.80	5.12	4.58
Animals Owned	676	34.47	16.43	11.84	8.44	9.92	7.55	4.89	3.70	2.82
Has. Land	223	27.80	18.39	10.76	10.31	8.52	4.93	7.17	8.52	3.59
Church Donations	197	31.47	23.86	11.68	6.09	16.24	5.58	2.03	1.02	2.03

Table A-3: χ^2 Tests and Kuiper V_N^* Tests between Benford's Law and Animals Owned in 1999 and 2002

	# of Surveys	# of Obs	χ^2	V_N^*
Total in 1999	298	827	56.21**	3.09**
Enumerator 1	33	105	10.73	0.88
Enumerator 2	36	103	8.18	1.09
Enumerator 3	33	91	14.87	1.39*
Enumerator 4	37	94	16.24*	1.80**
Enumerator 5	27	78	9.35	1.13
Enumerator 6	8	29	6.43	0.61
Enumerator 7	32	93	4.85	0.60
Enumerator 8	10	31	9.15	1.37*
Enumerator 9	32	84	14.86	1.33
Enumerator 10	22	50	20.44**	1.76**
Total in 2002	223	676	18.52*	1.14
Enumerator 1	71	211	7.58	0.83
Enumerator 2	75	223	15.81*	0.77
Enumerator 3	77	242	11.77	1.34

* indicates 95% and ** indicates 99% significantly different from Benford.

Table A-4: χ^2 Tests and Kuiper V_N^* Tests between Benford's Law and Quantities Produced in Paraguay

	# Surveys	# Obs	Obs per Survey	χ^2	V_N^*
Total in 1999	298	1412	4.7	30.36**	1.21
Enumerator 1	33	159	4.8	20.80**	1.57*
Enumerator 2	36	139	3.9	6.40	0.85
Enumerator 3	33	153	4.6	13.61	1.23
Enumerator 4	37	162	4.4	6.84	0.71
Enumerator 5	27	136	5.0	11.12	1.06
Enumerator 6	8	43	5.4	9.21	0.71
Enumerator 7	32	149	4.7	12.15	1.42*
Enumerator 8	10	60	6.0	7.03	0.44
Enumerator 9	32	202	6.3	14.33	1.47*
Enumerator 10	22	89	4.0	5.55	0.78
Total in 2002	223	1632	7.3	101.34**	2.69**
Enumerator 1	71	516	7.3	28.20**	1.50*
Enumerator 2	75	556	7.4	37.58**	1.68**
Enumerator 3	77	560	7.3	67.92**	1.97**
P.I. sat in	71	582	8.2	44.94**	1.47*
P.I. didn't sit in	152	1050	6.9	67.54**	2.26**

* indicates 95% and ** indicates 99% significantly different from Benford.

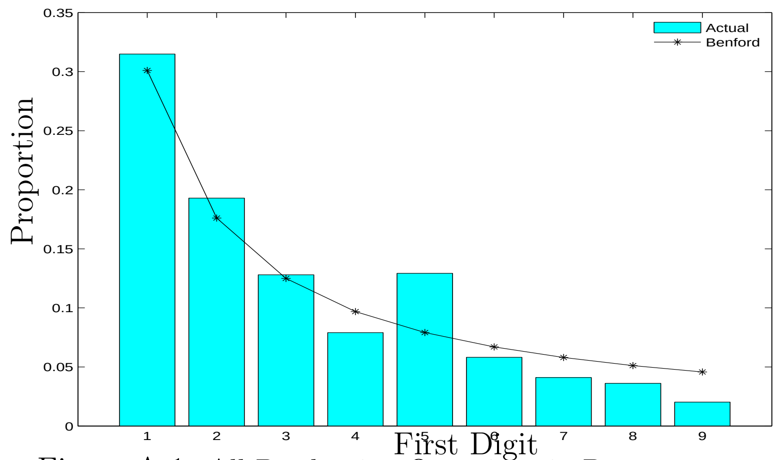

Figure A-1: All Production Quantities in Paraguay 2002

B Appendix: Published Papers Using Data Sets

B.1 Ghana - Markus Goldstein and Chris Udry

Conley, T. & Udry, C. (2001), 'Social learning through networks: The adoption of new agricultural technologies in Ghana', *American Journal of Agricultural Economics* **83**(3), 668–673.

Udry, C. & Anagol, S. (2006), 'The return to capital in Ghana', *The American Economic Review* **96**(2), 388–393.

B.2 Paraguay - Laura Schechter

Schechter, L. (2006), 'Traditional trust measurement and the risk confound: An experiment in rural Paraguay', *Journal of Economic Behavior and Organization* In Press.

B.3 Bangladesh - MHSS

Chaudhuri, A. (2005), 'Direct and indirect effects of a maternal and child health program in rural Bangladesh', *Journal of Developing Societies* **21**(1-2), 143–173.

Grira, H. (2004), 'The determinants of grade attainment in low-income countries: Evidence from rural Bangladesh', *The Developing Economies* **42**(4), 494–509.

Hoque, S. (2004), 'Micro-credit and the reduction of poverty in Bangladesh', *Journal of Contemporary Asia* **34**(1), 21–33.

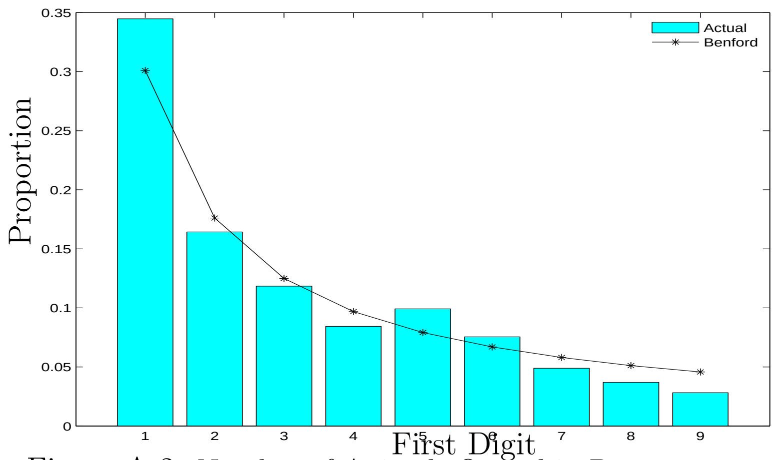

Figure A-2: Number of Animals Owned in Paraguay 2002

Kuhn, R. (2006), 'The effects of fathers' and siblings' migration on children's pace of schooling in rural Bangladesh', *Asian Population Studies* **2**(1), 69–92.

Maitra, P. (2003), 'Schooling and educational attainment: Evidence from Bangladesh', *Education Economics* **11**(2), 129–153.

Park, C. (2006), 'Risk pooling between households and risk-coping measures in developing countries: Evidence from rural Bangladesh', *Economic Development and Cultural Change* **54**(2), 423–457.

B.4 Mexico - PROGRESA

Behrman, J. R. & Hoddinott, J. (2005), 'Programme evaluation with unobserved heterogeneity and selective implementation: The Mexican PROGRESA impact on child nutrition', *Oxford Bulletin of Economics and Statistics* **67**(4), 547–569.

Behrman, J. R., Sengupta, P. & Todd, T. (2005), 'Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico', *Economic Development and Cultural Change* **54**(1), 237–275.

Coady, D. P. & Parker, S. W. (2004), 'Cost-effectiveness analysis of demand- and supply-side education interventions: The case of PROGRESA in Mexico', *Review of Development Economics* **8**(3), 440–451.

de Janvry, A., Finan, F., Sadoulet, E. & Vakis, R. (2006), 'Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks?', *Journal of Development Economics* **79**(2), 349–373.

Finan, F., Sadoulet, E. & de Janvry, A. (2005), 'Measuring the poverty reduction potential of land in rural Mexico', *Journal of Development Economics* **77**(1), 27–51.

Gertler, P. (2004), 'Do conditional cash transfers improve child health? Evidence from PROGRESA's Control Randomized Experiments', *American Economic Review* **94**(2), 336–341.

Hoddinott, J. & Skoufias E. (2004), ‘The impact of PROGRESA on food consumption’, *Economic Development and Cultural Change* **53**(1), 37–61.

Schultz, T. P. (2004), ‘School subsidies for the poor: Evaluating the Mexican PROGRESA poverty program’, *Journal of Development Economics* **74**(1), 199–250.

Skoufias, E., Davis, B. & de la Vega, S. ‘Targeting the poor in Mexico: An evaluation of the selection of households into PROGRESA’, *World Development* **29**(10), 1769–1784.

Skoufias, E. & Parker, S. W. (2001), ‘Conditional cash transfers and their impact on child work and schooling: Evidence from the PROGRESA program in Mexico’, *Economia* **2**(1), 45–96.

Stecklov, G., Winters, P., Stampini, M. & Davis, B. (2005) ‘Do conditional cash transfers influence migration? A study using experimental data from the Mexican PROGRESA program’, *Demography* **42**(4), 769–790.

B.5 Pakistan - IFPRI

Adams, R. H. (1995), ‘Agricultural income, cash crops, and inequality in rural Pakistan’, *Economic Development and Cultural Change* **43**(3), 467–491.

Adams, R. H. (1998), ‘Remittances, investment, and rural asset accumulation in Pakistan’, *Economic Development and Cultural Change* **47**(1), 155–173.

Alderman, H. (1996), ‘Saving and economic shocks in rural Pakistan’, *Journal of Development Economics* **51**(2), 343–365.

Alderman, H., Behrman, J. R., Khan, S., Ross, D. R. & Sabot, R. (1997), ‘The income gap in cognitive skills in rural Pakistan’, *Economic Development and Cultural Change* **46**(1), 97–122.

Alderman, H., Behrman, J. R., Lavy, V. & Menon, R. (2001), ‘Child health and school enrollment: A longitudinal analysis’, *The Journal of Human Resources* **36**(1), 185–205.

Fafchamps, M. & Quisumbing, A. R. (1999), ‘Human capital, productivity, and labor allocation in rural Pakistan’, *The Journal of Human Resources* **34**(2), 369–406.

Foster, A. D. & Rosenzweig, M. R. (2001), ‘Imperfect commitment, altruism, and the family: Evidence from transfer behavior in low-income rural areas’, *Review of Economics and Statistics* **83**(3), 389–407.

Kochar, A. (1999), ‘Evaluating familial support for the elderly: The intrahousehold allocation of medical expenditures in rural Pakistan’, *Economic Development and Cultural Change* **47**(3), 621–656.

Kochar, A. (2000), ‘Parental benefits from intergenerational coresidence: Empirical evidence from rural Pakistan’, *Journal of Political Economy* **108**(6), 1184–1209.

McCulloch, N. & Baulch, B. (2000), ‘Simulating the impact of policy upon chronic and transitory poverty in rural Pakistan’, *Journal of Development Studies* **36**(6), 100–130.

Ogaki, M. & Zhang, Q. (2001), ‘Decreasing relative risk aversion and tests of risk sharing’, *Econometrica* **69**(2), 515–526.

B.6 Peru - PLSS

Arriagada, A. M. (1990), 'Labor market outcomes of non-formal training for male and female workers in Peru', *Economics of Education Review* **9**(4), 331–342.

Bellew, R. & Moock, P. (1990), 'Vocational and technical education in Peru', *Economics of Education Review* **9**(4), 365–375.

Bhaumik, S. K. & Nugent, J. B. (1999), 'Analysis of food demand in Peru: Implications for food-feed competition', *Review of Development Economics* **3**(3), 242–257.

Bollen, K. A., Glanville, J. L. & Stecklov, G. (2002), 'Economic status proxies in studies of fertility in developing countries: Does the measure matter?', *Population Studies* **56**(1), 81–96.

Cox, D., Eser, Z. & Jimenez, E. (1998), 'Motives for private transfers over the life cycle: An analytical framework and evidence for Peru', *Journal of Development Economics* **55**(1), 57–80.

Cox, D. & Jimenez, E. (1992), 'Social security and private transfers in developing countries: The case of Peru', *World Bank Economic Review* **6**(1), 155–169.

Falaris, E. M. (2003), 'The effect of survey attrition in longitudinal surveys: Evidence from Peru, Cote d'Ivoire and Vietnam', *Journal of Development Economics* **70**(1), 133–157.

Gertler, P. & Glewwe, P. (1990), 'The willingness to pay for education in developing countries: Evidence from rural Peru', *Journal of Public Economics* **42**(3), 251–275.

Gertler, P. & Glewwe, P. (1992), 'The willingness to pay for education for daughters in contrast to sons: Evidence from rural Peru', *World Bank Economic Review* **6**(1), 171–188.

Glewwe, P. & de Tray, D. (1991), 'The poor in Latin America during adjustment: A case study of Peru', *Economic Development and Cultural Change* **40**(1), 27–54.

Glewwe, P. & Hall, G. (1994), 'Poverty, inequality, and living standards during unorthodox adjustment: The case of Peru, 1985-1990', *Economic Development and Cultural Change* **42**(4), 689–717.

Glewwe, P. & Hall, G. (1998), 'Are some groups more vulnerable to macroeconomic shocks than others? Hypothesis tests based on panel data from Peru', *Journal of Development Economics* **56**(1), 181–206.

Jacoby, H. G. (1991), 'Productivity of men and women and the sexual division of labor in peasant agriculture of the Peruvian Sierra', *Journal of Development Economics* **37**(1-2), 265–287.

Jacoby, H. G. (1993), 'Shadow wages and peasant family labour supply: An econometric application to the Peruvian Sierra', *Review of Economic Studies* **60**(4), 903–921.

Jacoby, H. G. (1994), 'Borrowing constraints and progress through school: Evidence from Peru', *The Review of Economics and Statistics* **76**(1), 151–160.

Lancaster, G. & Ray, R. (2002), 'International poverty comparisons on unit record data of developing and developed countries', *Australian Economic Papers* **41**(2), 129–139.

Lancaster, G., Ray, R. & Valenzuela, M. R. (1999), 'A cross-country study of equivalence scales and expenditure inequality on unit record household budget data', *Review of Income* *and Wealth* **45**(4), 455–482.

Lancaster, G., Ray, R. & Valenzuela, M. R. (1999), 'A cross-country study of household poverty and inequality on unit record household budget data', *Economic Development and Cultural Change* **48**(1), 177–208.

Laszlo, S. (2005), 'Self-employment earnings and returns to education in rural Peru', *Journal of Development Studies* **41**(7), 1247–1287.

MacIsaac, D. J. & Patrinos, H. A. (1995), 'Labour market discrimination against indigenous people in Peru', *Journal of Development Studies* **32**(2), 218–233.

Maitra, P. & Ray, R. (2002), 'The joint estimation of child participation in schooling and employment: Comparative evidence from three continents', *Oxford Development Studies* **30**(1), 41–62.

Marcouiller, D., de Castilla, V. R. & Woodruff, C. (1997), 'Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru', *Economic Development and Cultural Change* **45**(2), 367–392.

Montgomery, M. R., Gragnolati, M., Burke, K. A. & Paredes, E. (2000), 'Measuring living standards with proxy variables', *Demography* **37**(2), 155–174.

Newman, J. & Gertler, P. (1994), 'Family productivity, labor supply, and welfare in a low income country', *The Journal of Human Resources* **29**(4), 989–1026.

Pal, S. (2004), 'Child schooling in Peru: Evidence from a sequential analysis of school progression', *Journal of Population Economics* **17**(4), 657–680.

Patrinos, H. A. & Psacharopoulos, G. (1997), 'Family size, schooling and child labor in Peru: An empirical analysis', *Journal of Population Economics* **10**(4), 387–405.

Paxson, C. & Schady, N. (2005), 'Child health and economic crisis in Peru', *World Bank Economic Review* **19**(2), 203–223.

Pessino, C. (1991), 'Sequential migration theory and evidence from Peru', *Journal of Development Economics* **36**(1), 55–87.

Ray, R. (2000), 'Analysis of child labour in Peru and Pakistan: A comparative study', *Journal of Population Economics* **13**(1), 3–19.

Ray, R. (2000), 'Child labor, child schooling, and their interaction with adult labor: Empirical evidence for Peru and Pakistan', *The World Bank Economic Review* **14**(2), 347–367.

Schaffner, J. A. (1998), 'Premiums to employment in larger establishments: Evidence from Peru', *Journal of Development Economics* **55**(1), 81–113.

Schaffner, J. A. (2002), 'Heteroskedastic sample selection and developing-country wage equations', *Review of Economics and Statistics* **84**(2), 269–280.

Stelcner, M., van der Gaag, J. & Vijverberg, W. (1989), 'A switching regression model of public-private sector wage differentials in Peru, 1985-86', *The Journal of Human Resources* **24**(3), 545–559.

van der Gaag, J., Stelcner, M. & Vijverberg, W. (), 'Wage differentials and moonlighting by civil servants: Evidence from Cote d'Ivoire and Peru', *World Bank Economic Review* **3**(1), 67–95.

Yamada, G. (1996), 'Urban informal employment and self-employment in developing countries: Theory and evidence', *Economic Development and Cultural Change* **44**(2), 289–314.

Zwane, A. P. (2006), 'Does poverty constrain deforestation? Econometric evidence from Peru', *Journal of Development Economics* In Press.

B.7 Vietnam - VLSS

Deolalikar, A. B. (2002), 'Access to health services by the poor and the non-poor: The case of Vietnam', *Journal of Asian and African Studies* **37**(2), 244–261.

Falaris, E. M. (2003), 'The effect of survey attrition in longitudinal surveys: Evidence from Peru, Cote d'Ivoire and Vietnam', *Journal of Development Economics* **70**(1), 133–157.

Fritzen, S. (2002), 'Growth, inequality and the future of poverty reduction in Vietnam', *Journal of Asian Economics* **13**(5), 635–657.

Glewwe, P., Gragnolati, M. & Zaman, H. (2002), 'Who gained from Vietnam's boom in the 1990s?', *Economic Development and Cultural Change* **50**(4), 773–792.

Glewwe, P. & Jacoby, H. G. (2004), 'Economic growth and the demand for education: Is there a wealth effect?', *Journal of Development Economics* **74**(1), 33–51.

Haughton, D. & Haughton, J. (1997), 'Explaining child nutrition in Vietnam', *Economic Development and Cultural Change* **45**(3), 541–556.

Haughton, J. (1997), 'Falling fertility in Vietnam', *Population Studies* **51**(2), 203–211.

Haughton, J. & Haughton, D. (1995), 'Son preference in Vietnam', *Studies in Family Planning* **26**(6), 325–337 .

Minot, N. & Goletti, F. (1998), 'Export liberalization and household welfare: The case of rice in Vietnam', *American Journal of Agricultural Economics* **80**(4), 738–749.

Nguyen, B. T., Albrecht, J. W., Vroman, S. B. & Westbrook, D. M. (2006), 'A quantile regression decomposition of urban-rural inequality in Vietnam', *Journal of Development Economics* In Press.

Ravallion, M. & van de Walle, D. (2004), 'Breaking up the collective farms', *Economics of Transition* **12**(2), 201–236.

Sepehri, A., Chernomas, R. & Akram-Lodhi, A. H. (2003), 'If they get sick, they are in trouble: Health care restructuring, user charges, and equity in Vietnam', *International Journal of Health Services* **33**(1), 137–161.

van de Walle, D. (2003), 'Are returns to investment lower for the poor? Human and physical capital interactions in rural Vietnam', *Review of Development Economics* **7**(4), 636–653.

van de Walle, D. & Cratty, D. (2004), 'Is the emerging non-farm market economy the route out of poverty in Vietnam?', *Economics of Transition* **12**(2), 237–274.

van de Walle, D. & Gunewardena, D. (2001), 'Sources of ethnic inequality in Vietnam', *Journal of Development Economics* **65**(1), 177–207.

References

Benford, F. (1938), 'The law of anomalous numbers', *Proceedings of the American Philosophical Society* **78**(4), 551–572.

Berger, A. & Hill, T. P. (2006), 'Newton's method obeys Benford's law', *American Mathematical Monthly* . Forthcoming.

Carslaw, C. A. P. N. (1988), 'Anomalies in income numbers: Evidence of goal oriented behavior', *Accounting Review* **63**(2), 321–327.

de Marchi, S. & Hamilton, J. T. (2006), 'Assessing the accuracy of self-reported data: An evaluation of the toxics release inventory', *Journal of Risk and Uncertainty* **32**, 57–76.

Deaton, A. (2005), 'Measuring poverty in a growing world (or measuring growth in a poor world)', *The Review of Economics and Statistics* **87**(1), 1–19.

Deming, W. E. (2006), 'On errors in surveys (an excerpt)', *The American Statistician* **60**(1), 34–38.

Diaconis, P. (1977), 'The distribution of leading digits and uniform distribution mod 1', *The Annals of Probability* **5**(1), 72–81.

Durtschi, C., Hillison, W. & Pacini, C. (2004), 'The effective use of Benford's law to assist in detecting fraud in accounting data', *Journal of Forensic Accounting* **5**, 17–34.

Geyer, C. L. & Williamson, P. P. (2004), 'Detecting fraud in data sets using Benford's law', *Computation in Statistics: Simulation and Computation* **33**(1), 229–246.

Giles, D. E. (2006), 'Benford's law and naturally occurring prices in certain ebaY auctions', *Applied Economics Letters* . Forthcoming.

Glewwe, P. & Dang, H.-A. H. (2005), The impact of decentralized data entry on the quality of household survey data in developing countries: Evidence from a randomized experiment in Vietnam. University of Minnesota Working Paper.

Grendar, M., Judge, G. & Schechter, L. (2006), An empirical non-parametric likelihood family of data-based Benford-like distributions. Unpublished Manuscript.

Groves, R. M. (1989), *Survey Errors and Survey Costs*, New York: John Wiley & Sons.

Hill, T. P. (1995), 'A statistical derivation of the significant-digit law', *Statistical Science* **10**(4), 354–363.

Hill, T. P. (1998), 'The first digit phenomenon', *American Scientist* **86**(4), 358–363.

Hill, T. P. & Schürger, K. (2005), 'Regularity of digits and significant digits of random variables', *Journal of Stochastic Processes and Their Applications* **115**, 1723–1743.

Leemis, L. M., Schmeiser, B. W. & Evans, D. L. (2000), 'Survival distributions satisfying Benford's law', *The American Statistician* **54**(4), 236–241.

Livio, M. (2002), *The Golden Ratio: The Story of Phi, the World's Most Astonishing Number*, Broadway, New York, NY.

Morrow, J. (2006), Benford's law and families of distributions. Unpublished Manuscript.

Newcomb, S. (1881), 'Note on the frequency of use of the different digits in natural numbers', *American Journal of Mathematics* **4**, 39–40.

Nigrini, M. J. (1996), 'A taxpayer compliance application of Benford's law', *Journal of the American Taxation Association* **18**(1), 72–91.

Nigrini, M. J. (1999), 'Adding value with digital analysis', *The Internal Auditor* **56**(1), 21–23.

Nigrini, M. J. & Miller, S. J. (2006), Benford's law applied to hydrology data: Results and relevance to other geophysical data. Working Paper, Brown University.

Philipson, T. & Malani, A. (1999), 'Measurement errors: A principal investigator-agent approach', *Journal of Econometrics* **91**(2), 273–298.

Raimi, R. (1976), 'The first digit problem', *American Mathematical Monthly* **83**, 521–538.

Rodriguez, R. J. (2004), 'First significant digit patterns from mixtures of uniform distributions', *The American Statistician* **58**(1), 64–71.

Schatte, P. (1988), 'On mantissa distributions in computing and Benford's law', *Journal of Information Processing and Cybernetics* **24**(10), 443–455.

Scott, P. D. & Fasli, M. (2001), 'Benford's law: An empirical investigation and a novel explanation'. University of Essex Working Paper.

Stephens, M. A. (1970), 'Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables', *Journal of the Royal Statistical Society, Series B* **32**(1), 115–122.

Thomas, D., Frankenberg, E. & Smith, J. P. (2001), 'Lost but not forgotten', *The Journal of Human Resources* **36**(3), 556–592.

Varian, H. (1972), 'Benford's law', *The American Statistician* **26**, 65.