Do the Rich Get Richer in the Stock Market? Evidence from India

John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish[1]

First draft: March 2018

[1]Campbell: Department of Economics, Littauer Center, Harvard University, Cambridge MA 02138, USA, and NBER. john_campbell@harvard.edu. Ramadorai: Imperial College London, London SW7 2AZ, UK. t.ramadorai@imperial.ac.uk. Ranish: Federal Reserve Board, 20th & C Streets NW, Washington DC 20551, USA. ben.ranish@frb.gov. We gratefully acknowledge NSDL and SEBI for providing us with access to the data, the Sloan Foundation for financial support, and Imperial College for the use of computing facilities. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Board of Governors or other employees of the Federal Reserve System.

Abstract

We use data on Indian stock portfolios to show that return heterogeneity is the primary contributor to increasing inequality of wealth held in risky assets by Indian individual investors. Return heterogeneity increases wealth inequality through two main channels, both of which are related to the prevalence of undiversified accounts that own relatively few stocks. First, some undiversified portfolios randomly do well, while others randomly do poorly. Second, larger accounts diversify more effectively and thereby earn higher average log returns even though their average simple returns are no higher than those of smaller accounts.

1 Introduction

New methods for imputing the wealth distribution have provided evidence that wealth inequality is increasing both in the US and globally (Alvaredo et al 2018, Cagetti and De Nardi 2008, Saez and Zucman 2016). Wealth inequality results in part from income inequality, but it may also be driven by bequests and by unequal returns earned on financial investments, particularly if returns are higher for those who are already wealthier. Piketty (2014, Ch. 12) emphasizes the last factor when he writes: "Many economic models assume that the return on capital is the same for all owners, no matter how large or small their fortunes. This is far from certain, however: it is perfectly possible that wealthier people obtain higher average returns than less wealthy people." The purpose of this paper is to measure the contribution of heterogeneous investment returns to wealth inequality using detailed administrative data on the equity portfolios of Indian stock market investors.

Investment returns multiply initial wealth; equivalently, log returns have an additive effect on initial log wealth. What matters for the evolution of wealth inequality is therefore heterogeneity in log returns and the correlation of log returns with initial log wealth. The finance literature has long recognized that the portfolio with the highest average log return will tend to have a growing share of wealth over time in the absence of inflows and outflows.[2] However, the literature on inequality has not always made the distinction between simple returns and log returns, a distinction that we show to be important for understanding the effect of return heterogeneity on wealth inequality.

The average log return on a risky portfolio is always less than the average simple return by Jensen's inequality. If the portfolio return is lognormally distributed, the average log return is the average simple return less one-half the variance of the log return.[3] Importantly, this implies that investors can earn a higher average log return not only by earning a higher average simple return, but also by diversifying more effectively, thereby lowering portfolio

[2] This "growth-optimal" portfolio will be chosen by a rational investor with log utility. It outperforms any other portfolio with increasing probability as the investment horizon increases (Markowitz 1976).

[3] If returns are not lognormally distributed, then the difference between average simple return and average log return is the entropy of the return, a more general measure of dispersion that involves higher moments as well as variance. Campbell (2018, p.100) provides a textbook exposition.

variance with an unchanged average simple return. As a stark example, in a market with many stocks whose returns are identically distributed and imperfectly correlated, all portfolios have the same average simple return, but better diversified portfolios that hold more stocks have lower variances and higher average log returns.

This analysis implies that heterogeneous returns can contribute to wealth inequality through several channels. First, undiversified risktaking causes random cross-sectional variation in realized log returns: in each period, some undiversified investors get lucky while others are unlucky, and this causes wealth levels to diverge. Second, cross-sectional variation in the degree of undiversified risktaking lowers the average log returns on less diversified, more volatile portfolios relative to their average simple returns. Third, there can be cross-sectional variation in average simple returns resulting from heterogeneity in investors' willingness to take risk, ability to identify compensated risk exposures, or stock-picking skill in an inefficient market.[4] The second and third factors are particularly important if they correlate with initial wealth levels.

Measuring investment returns is a challenging task that requires even more data than measuring wealth inequality. The latter requires snapshots of portfolio values at points in time, while the former requires in addition either detailed knowledge of portfolio composition and of individual asset returns during the intervals between snapshots, or a complete time series of portfolio inflows and outflows that can be used to impute returns. In this paper we work with data on directly held Indian equities, whose ownership is electronically recorded and linked to over ten million equity accounts held by Indian individual investors. These data enable us to accurately measure the returns that investors earn in the public equity market and hence quantify the contribution of heterogeneous returns on directly held stocks to inequality in the size of equity accounts. We create a random sample of 200,000 accounts and measure inequality using the cross-sectional variance of log account size (the log market value of equities held) which relates cleanly to the properties of log returns discussed above.

During our sample period, March 2002 to May 2011, the cross-sectional variance of log

[4] Although outside the scope of this paper, average simple returns may also be unequal if some investors have privileged access to certain asset classes such as private equity or real estate.

account size increased by about two-thirds, from 3.4 to 5.7: equivalently, a one-standard deviation increase in log account size multiplied market value by 6.4 at the start of the sample and by 10.8 at the end. We break this increase into four components, which reflect heterogeneous log returns, heterogeneous net inflows, the covariance between returns and inflows, and account entry and exit, respectively. The heterogeneous log return component is 84% of the total, while heterogeneous flows contribute 40%, the covariance between returns and flows −23%, and account entry and exit −1%.

We further decompose the contribution of return heterogeneity using a framework proposed by Campbell (2016). We show that the first two channels discussed above are both operative. Heterogeneous log returns contribute to wealth inequality both through the random realizations of underdiversified portfolio returns, and through higher average log returns that wealthier investors earn on their equity investments. Importantly, the third channel is not operative in our data because wealthier investors do not have higher average simple returns. In fact, the opposite is true, because smaller investors have higher loadings on compensated risk factors including the market, small-stock, and value factors.[5] Rather, wealthier investors are better diversified so their idiosyncratic uncompensated risk is lower, and they earn higher Sharpe ratios and higher average log returns.

The Indian context of our empirical study is of interest not only because it provides us with complete and accurate data on portfolio composition for directly held stocks, but also because India is a large country with a rapidly growing middle class that is starting to use risky financial markets for the first time, and that during our sample period invests in stocks directly rather than through mutual funds. In these respects India has much in common with other developing countries. The contributors to wealth inequality in this context are likely different than they are in a developed market with a high mutual fund share.

There is little research on return heterogeneity in the US because of data limitations, but several papers have documented return heterogeneity in Sweden, for example Calvet,

[5]There is suggestive evidence that wealthier investors have stock-picking skill relative to the standard four-factor asset pricing model, due to Fama and French (1993) and Carhart (1997), that we use to model compensated risk. However, this is not sufficient to offset the lower factor risk loadings in wealthier investors' portfolios.

Campbell, and Sodini (2007) and Bach, Calvet, and Sodini (BCS 2015).[6] These papers find that Swedish investors are relatively well diversified, so the major contributor to return heterogeneity in Sweden is the willingness of richer investors to earn higher returns by taking more equity risk. This phenomenon does not contribute to our results because we observe only equity portfolios and not holdings of safe assets.

The outline of the paper is as follows. Section 2 describes our data. Section 3 discusses the variation in risk and return by account size. Section 4 decomposes the change in the variance of wealth inequality across Indian equity accounts. Section 5 summarizes robustness exercises presented in our online appendix (Campbell, Ramadorai, and Ranish 2018), and section 6 concludes.

2 Data

Our data on Indian equity accounts are described in detail in Campbell, Ramadorai, and Ranish (CRR 2014). They come from India's National Securities Depository Limited (NSDL), with the approval of the Securities and Exchange Board of India (SEBI). NSDL is the larger of two securities depositories in India, with a roughly 80% market share of total assets tracked and a 60% market share by number of accounts. During our sample period almost all equities held and almost all transactions were recorded electronically.

These data do have a few limitations that should be noted. First, we have little information about account holders beyond a type classification, which we use to separate Indian individual investors from others including beneficial owners, domestic financial and non-financial institutions, foreign investors, and government accounts.

Second, we do not observe individual investors' holdings of mutual funds. This is not a major omission because during our sample period the fraction of equity market capitalization held by mutual funds was modest in India, always less than 5%. In addition, roughly 60% of mutual funds in India are held by corporations. CRR estimate that individuals' indirect equity holdings through mutual funds and related products were between 6% and 19% of

[6] Fagereng et al. (2016) present a similarly comprehensive analysis of return heterogeneity in Norway.

total household equity holdings over the sample period. They also note that a 2009 SEBI survey found that about 65% of Indian households owning individual stocks did not own any bonds or mutual funds.

Third, we do not observe data on the derivatives transactions of Indian investors, including their participation in single-stock futures markets. However, while single-stock futures volume is considerable in India, larger in fact than equity index futures volume (Martins, Singh, and Bhattacharya 2012, Vashishtha and Kumar 2010), it is likely concentrated in a small minority of accounts and unimportant for the majority of Indian investors.

A single investor can hold multiple accounts on NSDL; however, we link these together using each investor's Permanent Account Number (PAN), a unique taxpayer identifier. PAN aggregation reduces the total number of individual accounts in our database from about 13.7 million to 11.6 million.

The fraction of Indian equity market capitalization that is held in NSDL accounts grows from just above 50% in 2002 to about 70% in 2011. The share of this held in individual accounts declines from about 20% to about 10%, reflecting changes in NSDL coverage of institutions as well as an increase in institutional investment. The number of individuals holding stock in NSDL accounts grows from 2.28 to 6.25 million, that is, by about 175%.[7]

We obtain monthly data on stock returns from Prowess, Datastream, and Compustat Global. In addition, we impute price returns from our NSDL data. We use only those returns that we are able to validate through comparison between at least two of the data sources. (We follow a similar approach to validating stocks' book-market ratios and market capitalization.) In addition, we both attempt to manually fill otherwise missing returns for the few instances where a stock with a missing return comprises at least 1% of the average individual's stock portfolio, and manually validate the 25 largest and smallest percentage returns. Overall, we use returns that on average cover slightly more than 95% of an individual account's stock holdings.[8]

[7] Because of account exit, the number of individuals holding stock at any point in time is always considerably smaller than the total number of individual stockholders in our sample.

[8] We compute account-level returns using only those stocks for which we have validated returns. In our variance decomposition, changes in value due to missing returns are captured by the "net inflows" component

2.1 Summary statistics

Our results are estimated from a sample of 200,000 accounts selected randomly from accounts that held stock at any time during our sample period. Table 1 presents summary statistics on this sample, reporting the time-series average, minimum, maximum, and standard deviation for a series of cross-sectional statistics calculated at the end of each month from March 2002 to May 2011. The number of stockholding accounts in the sample varies from about 39,000 to about 108,000, with an average of 74,000. The time-series average account entry rate is 2.8% per month, and the exit rate 1.9% per month, but the entrance rate in particular is highly variable over time as IPOs and high returns attracted many Indian investors to begin participating in the stock market during the mid-2000s.[9]

The cross-sectional mean log account size varied during the period from 10.32 to 11.53, corresponding to about 30,000 and 100,000 rupees respectively, or $660 and $2,211 using the sample average exchange rate of 46 rupees per dollar. On average across all months, the cross-sectional mean growth rate of account size was about 1.7% per month, very similar to the cross-sectional mean log return of 1.4% on investments in these accounts. The difference between these two numbers, the average contribution of net inflows to mean account growth, was small at 29 basis points.

The main focus of our paper is on changes over time in account size inequality. Table 1 reports that the cross-sectional standard deviation of log account size increased from 1.85 to 2.38 during our sample period, corresponding to variances of 3.4 and 5.7 respectively. At the beginning of our sample, a one-standard deviation increase in log account size multiplied account size by 6.4, whereas at the end it multiplied account size by 10.8. Figure 1 reports the probability density function (panel A) and cumulative distribution function (panel B) of account size in the first and last months of our sample. The increase in variance is easily

provided that at least some returns are available for the account. In about 2.5% of account-months, we are missing returns for all stocks held. These account months are excluded from our risk and return analyses, and contribute to the "entrance and exit" component of our variance decomposition.

[9] Some accounts enter and exit multiple times as investors hold stocks, divest them, and subsequently acquire stocks again. Table 1 reports the average "first-time" entrance rate as 2.0%, implying that the average re-entrance rate is 0.8%.

visible, as is a spike in the probability density (a steep section of the CDF) around \$70 in the last month of the sample. This is the result of IPO subscriptions which were allocated in standard amounts to many small accounts that hold undiversified single-IPO positions. Converting account sizes to US dollars, the 10th percentile of account size fell during our sample period from \$71 to \$60, while the 90th percentile increased from \$7,274 to \$19,258.

The remainder of the paper asks what forces contribute to this increasing inequality in account size. We will show that the dominant influence on the evolution of account size inequality is the heterogeneity of investment returns. This is true despite the high cross-sectional volatility of net inflows reported in Table 1.[10] While net inflows are volatile, they are also negatively correlated with log account size, which greatly reduces their influence on the evolution of inequality.

3 Risk and Return by Account Size

In this section we examine variation in portfolio characteristics by account size. In each month we divide accounts into deciles by their value at the end of the previous month, equally weight accounts within each decile, and report summary statistics by decile.

Panel A of Table 2 reports arithmetic average excess returns by decile, from the smallest at the left to the largest at the right. The far right hand column of the table reports the difference in excess returns between the largest and the smallest decile accounts. The smallest accounts earn an excess return of 2.99% per month in this sample period, while the largest earn only 1.70% per month. The difference in excess returns between the two extreme portfolios is −1.29%, but the estimate is noisy with a standard error of 0.78%. The large standard error reflects our short sample period of less than ten years, the volatility of Indian stock returns, and systematic differences in the investment styles of large and small investors.

[10] The time-series average of the cross-sectional standard deviation of net inflows is 40.5%, almost as large as the cross-sectional standard deviation of account growth at 41.1%. The time-series average of the cross-sectional standard deviation of log returns is smaller at 9.1%.

Table 2 uses a standard four-factor model, due to Fama and French (1993) and extended by Carhart (1997), to measure size patterns in investment styles. The table shows that the smallest account returns have a beta with the market index of 1.21, while the largest have a beta of 1.00, and the difference of -0.21 has a standard error of 0.05. The smallest accounts load strongly on the size or SMB (small minus big) factor, with a loading of 0.57 as compared to 0.06 for the largest accounts; the difference of -0.51 has a standard error of 0.05. The smallest accounts load strongly on the value or HML (high minus low book-to-market) factor, with a loading of 0.45 as compared to -0.07 for the largest accounts; the difference of -0.51 has a standard error of 0.16. All three of these factors have positive average returns, both in India in our sample period and globally over much longer periods of time. Hence, these three factor loadings contribute to the higher average returns earned by smaller Indian investors. However, the smallest accounts have a negative loading of -0.36 on Carhart's (1997) momentum or MOM factor, while the largest accounts have a much smaller negative loading of -0.07; the difference of 0.29 has a standard error of 0.07. Since momentum also has a positive average return, this is the one factor loading that should deliver higher average returns to larger accounts.

As a reality check, Panel B of Table 2 relates these factor loadings to the average characteristics of the stocks held by different sizes of accounts. The patterns in factor loadings show up clearly in stock characteristics: the smallest accounts hold stocks with higher predicted betas, lower market capitalization (extreme among the smallest accounts), higher book-to-market ratios, and lower realized returns over the previous year excluding the previous month.

The market, size, and value tilts reported in Table 2 for small Indian investors are similar to those reported in Barber and Odean (2000, Table II) for a sample of US retail investors. However, the effects of account size on these tilts differ from those reported by BCS (2015) for the cross-section of Swedish investors. BCS find that in Sweden, wealthier investors have higher loadings on market, size, and value factors than poorer investors do. One reason for the difference in results is likely that poorer Swedish investors tend to hold mutual funds with minimal style tilts. We are unaware of evidence on cross-sectional variation in momentum tilts among retail investors, but Kaniel, Saar, and Titman (2008) report that retail investors as a group tend to be contrarians which is consistent with our findings.

Panel A of Table 2 also reports alphas from the four-factor model, that is, the components of average excess returns not explained by factor loadings and average excess returns to the factors. The smallest accounts have a negative alpha of -0.17% per month, while the largest accounts have a positive alpha of 0.40% per month. These point estimates suggest that larger Indian investors have stock-picking skill relative to the four-factor Fama-French (1993) and Carhart (1997) model. However, the alpha spread of 0.57% has a large standard error of 0.46%, again unsurprising given our relatively short sample period.

We have emphasized that average log returns depend not only on average simple returns, but also on diversification. Panel B of Table 2 reports the average number of stocks held in each decile of account size. This increases strongly from 1.6 in the smallest decile to 28.9 in the largest decile, so large Indian equity accounts are far better diversified than small ones.

One would expect this size pattern in diversification to show up in the volatilities of account returns, and we examine this in Table 3. The first row of Table 3 shows that small Indian investors hold highly volatile portfolios, with an average standard deviation of 23.7% per month (equivalent to 82% per year). The portfolios of the largest Indian investors have a much lower standard deviation of 11.0% per month (38% per year), and the difference of -12.7% has a standard error of only 0.9%. The second row of Table 3 repeats the average excess returns from Table 2. The third row of Table 3 takes the ratio of the average excess return to standard deviation to calculate the Sharpe ratio. This is somewhat lower for the smallest accounts at 0.13 than for the largest accounts at 0.15, although the difference of 0.03 has a standard error of 0.04.

The fourth row of Table 3 calculates an average excess log return. Unlike the average excess simple return, the average excess log return is increasing in account value, 0.72% per month for the smallest accounts and 1.10% per month for the largest accounts, although the difference of 0.38% has a standard error of 0.69%. The reason why average excess log return increases with account value is that volatility lowers average log return for any given level of average simple return. In the Indian data, the lower volatility of large accounts outweighs their lower average simple returns and gives them a higher average log return.

3.1 Correcting for luck

One reason why smaller Indian investors enjoyed high average simple returns during our sample period is that their style tilts performed spectacularly well. The average excess return on the Indian market was 1.46% per month in this period, while in global data from November 1990 through November 2017, the average excess return on a global index was only 0.53% per month. Similarly, the average returns on SMB and HML were 0.91% per month and 2.49% per month in India during our sample, but only 0.06% and 0.33% per month in the longer run global data. The fourth factor, MOM, delivered 0.64% per month in India, which is barely above the longer run global average of 0.60%, but this factor was favored by larger rather than by smaller investors. In other words, it is possible that small investors were lucky in this short sample period, and enjoyed higher returns than would normally be expected.

As a simple way to correct for this, the lower panels of Table 3 present counterfactual average excess returns, Sharpe ratios, and average excess log returns that would have been realized in our sample period if the four factors had delivered their long-run global average excess returns. This may be a more reasonable estimate of the returns that could have been expected on Indian equity portfolios ex ante. The middle panel preserves the alpha estimates from our primary analysis, and the bottom panel sets alpha to zero for all Indian investors. In the middle panel, expected excess returns are increasing in account value and in the bottom panel they are almost flat. In either case, Sharpe ratios and excess log returns are increasing in account size. The difference in excess log returns between the largest and smallest accounts is 2.15% in the first case, with a standard error of 0.46%, and 1.57% in the second case, with a standard error of 0.20%.

4 Decomposition of the Change in Wealth Inequality

In this section we ask how the patterns of risk and return we have documented affect the evolution of inequality in the account sizes of Indian equity investors. We use a simple accounting framework, an extension of one proposed by Campbell (2016).

Denote the market value of investor i's equity account at time t by W_{it}, and the gross return from t to $t+1$ on the account's time t investments by $(1+R_{i,t+1})$. For any account that exists in our data at both time t and time $t+1$, we can write

$$\begin{aligned} W_{i,t+1} &= W^0_{i,t+1} + F_{i,t+1} \\ &= W_{it}(1+R_{i,t+1})\left(1+\frac{F_{i,t+1}}{W^0_{i,t+1}}\right), \end{aligned} \tag{1}$$

where $W^0_{i,t+1} = W_{it}(1+R_{i,t+1})$ denotes the value of the account at time $t+1$ if the stocks held at time t are held over the full month with no other account activity, and $F_{i,t+1}$ captures the effect of intramonthly portfolio rebalancing, inflows, and outflows. In the simplest case where there is no trading in the portfolio except at the end of each month, and where inflows arrive immediately before account value is measured, $F_{i,t+1}$ is the net inflow at time $t+1$.

Taking logs, we have

$$w_{i,t+1} = w_{it} + r_{i,t+1} + f_{i,t+1}, \tag{2}$$

where $w_{it} = \log(W_{it})$, $r_{i,t+1} = \log(1+R_{i,t+1})$, and $f_{i,t+1} = \log(1+F_{i,t+1}/W^0_{i,t+1})$.

At each point in time we can calculate the cross-sectional variances and covariances of log account size, returns, and net inflows. We use the notation Var^*_t and Cov^*_t to denote these cross-sectional second moments. Then from equation (2), but allowing for account entry and exit to affect the cross-sectional distribution of account size, we have

$$\begin{aligned} \text{Var}^*_t(w_{i,t+1}) - \text{Var}^*_t(w_{it}) = {} & \text{Var}^*_t(r_{i,t+1}) + 2\text{Cov}^*_t(w_{it}, r_{i,t+1}) \\ & + \text{Var}^*_t(f_{i,t+1}) + 2\text{Cov}^*_t(w_{it,}, f_{i,t+1}) \\ & + 2\text{Cov}^*_t(r_{i,t+1,}, f_{i,t+1}) + x_{i,t+1}. \end{aligned} \tag{3}$$

The first two terms on the right hand side of equation (3) are the contribution of log return inequality to the change in log account size inequality; the next two terms are the contribution of net inflow inequality; the fifth term is an interaction effect between the two; and the last term $x_{i,t+1}$ is a residual that captures the effects of account entry and exit. If we confined attention to accounts that exist both at time t and at time $t+1$, then $x_{i,t+1}$ would be zero.

Panel A of Table 4 presents the time-series average values of these terms in our Indian data. The average change in the cross-sectional variance of log wealth is 0.0197 per month. The contribution of log return inequality is 0.0166 or 84% of the total; the contribution of flow inequality is 0.0079 or 40%; the interaction effect is −0.0046 or −23%; and the effect of account entry and exit is a modest −0.0003 or −1% of the total. Thus, the dominant contributor to the increase in the inequality of account size in our data is indeed the heterogeneity in log investment returns.

4.1 Decomposition of the return contribution

We can go further in characterizing the contribution of log return inequality to account size inequality. Consider a model of the conditional expected log return on account i, where the expectation is formed at time t and applies to returns that are realized at time $t+1$. Write this conditional expected log return as μ_{it}. Then

$$r_{i,t+1} = \mu_{it} + \varepsilon_{i,t+1}, \tag{4}$$

where $\varepsilon_{i,t+1}$ is the unexpected return on account i at time $t+1$.

The contribution of log return inequality to account size inequality can be decomposed as

$$\begin{aligned}\mathrm{Var}_t^*(r_{i,t+1}) + 2\mathrm{Cov}_t^*(w_{it}, r_{i,t+1}) = {} & \mathrm{Var}_t^*(\mu_{it}) + \mathrm{Var}_t^*(\varepsilon_{i,t+1}) + 2\mathrm{Cov}_t^*(\mu_{it}, \varepsilon_{i,t+1}) \\ & + 2\mathrm{Cov}_t^*(w_{it}, \mu_{it}) + 2\mathrm{Cov}_t^*(w_{it}, \varepsilon_{i,t+1}). \end{aligned} \tag{5}$$

The first term on the right hand side of equation (5) is the cross-sectional variance of expected log returns, the result of heterogeneous investment strategies that offer different average log returns. The second term is the cross-sectional variance of unexpected log returns, the result of underdiversification. The third term is the covariance between expected and unexpected log returns; this can be nonzero in any cross-section, but should be zero if one takes a time-series average of equation (5) with a long enough sample period and if μ_{it} is a rational expectation of log return. The fourth term is the covariance between log wealth and expected log return; this captures the tendency of richer people to invest more effectively and earn higher log returns. The fifth term is the covariance between log wealth and unexpected log return; this can be nonzero in any cross-section, if investment strategies favored by wealthy people do better or worse than average, but should average to zero in a long sample period if μ_{it} is a rational expectation.

Panel B of Table 4 presents an empirical implementation of this decomposition. We consider three alternative models of conditional expected log returns. In model 1, the expected log return is simply the sample average return on accounts in the given size decile. In model 2, it is the counterfactual average return that would have been realized, if style portfolios had delivered their long-run global average returns rather than the extremely high returns realized in India during this period. However, model 2 uses our empirical estimates of four-factor alphas for Indian investors. Model 3 also uses long-run global average factor returns but sets the four-factor alphas to zero. These three models correspond to the three panels of Table 3.

In all three models, the cross-sectional variance of expected log returns has a negligible effect on the evolution of account size inequality, and the cross-sectional variance of unexpected log returns contributes about half of the total effect. The remaining half is attributed to the covariance between log wealth and realized log return.

Where the models differ is in the breakdown of this second half into the covariance of log wealth with expected log return and the covariance with unexpected log return. In model 1, sample average size decile returns are used and so the covariance with unexpected log return is close to zero (but not exactly zero because of small differences in average returns across accounts within each size decile). In other words, in model 1 the systematic ability of large investors to earn a higher average log return accounts for about half of the observed increase in wealth inequality.

In models 2 and 3, by contrast, long-run global average factor returns are used. Given the style tilts documented in Table 2, this implies that the covariance between log wealth and expected log return is even larger and the covariance between log wealth and unexpected log return is negative. According to these models, large investors have an even greater systematic ability to earn higher average log returns; this ability would have further increased wealth inequality in the Indian data, if it were not for the fact that smaller Indian investors "got lucky" by betting on factors that happened to outperform during this short sample period. Depending on the assumption about alpha, the effect of smaller investors' luck was to dampen the increase in wealth inequality by 36% to 75% of the total observed increase in log wealth inequality in this period.

4.2 Decomposition of the flow contribution

Panel C of Table 4 looks more carefully at the contribution of flow inequality to the evolution of wealth inequality. It may be surprising that the flow contribution is so modest in Panel A of Table 4 when the cross-sectional volatility of flows is so large in Table 1. The explanation is given in Panel C. Here we see that the covariance between log account size and flows is strongly negative, and it almost exactly cancels the contribution from the variance of flows. In other words, while flows are volatile, small accounts tend to have inflows while larger accounts have outflows (to fund spending or acquisition of other assets), and this limits the contribution of flows to the evolution of account size inequality. This phenomenon may be related to the negative covariance between returns and flows reported in the third row of Panel A.

5 Robustness

In the online appendix, we show that our results are robust to several variations in our empirical methodology. Other measures of wealth inequality also trend upwards in our sample period. Focusing on the right tail of the size distribution, we examine deciles within the top 5% of accounts and find fairly flat patterns of risk and return within this group. We control for account age and find similar results for cohort-balanced size deciles, each of which contains the same age distribution of accounts at each point in time. We exclude micro-cap stocks whose returns may be biased upwards by survivorship bias or bid-ask bounce (Blume and Stambaugh 1983) and again find similar results.

We have used accounts' market values to measure their size even though returns affect subsequent market values. Could it be that large accounts earn higher average log returns because there is persistent cross-sectional variation in investment skill, and skillful investors have accounts that become large? To address this concern, we construct two alternative measures of account size that are unaffected by realized returns, and find similar results when we sort accounts using these size measures.

Finally, we distinguish two types of account entry and exit. Accounts can enter or exit either because they start or cease to hold stocks, or because the stocks they hold start or cease to have at least one measured return. We show that entry and exit driven by stockholdings increases wealth inequality, while entry and exit driven by the availability of returns data decreases it. The two effects offset each other to create the modest −1% contribution of account entry and exit reported in Panel A of Table 4.

6 Conclusion

We have studied wealth held in equity accounts in India, a large developing country that is important for the evolution of global wealth inequality. We have shown that heterogeneous risky log investment returns have important effects on the cross-sectional distribution of account size: large accounts result not only from large contributions, but also from high log returns. The effect of log return heterogeneity accounts for 84% of the increase in the cross-sectional variance of log account size during our sample period from March 2002 to May 2011.

Return heterogeneity increases wealth inequality through two main channels, both of which are related to the prevalence of undiversified accounts that own relatively few stocks. The first is that some undiversified portfolios randomly do well, while others randomly do poorly. The second is that larger accounts tend to earn higher average log returns. They do so not by earning higher average simple returns, but by limiting uncompensated idiosyncratic risk which lowers the average log return for any given average simple return. Thus our paper highlights the importance of investment vehicles, such as mutual funds and exchange traded funds, that allow small investors to diversify risk; and it partially supports Piketty's (2014) concern that the rich get richer by earning high investment returns, subject to the distinction, central in finance theory, between simple and log returns.

References

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. *World Inequality Report 2018.* World Inequality Lab.

Bach, Laurent, Laurent Calvet, and Paolo Sodini. 2015. "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." Unpublished paper.

Barber, Brad M., and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55: 773–806.

Blume, Marshall E., and Robert F. Stambaugh. 1983. "Biases in Computed Returns: An Application to the Size Effect." *Journal of Financial Economics* 12: 387–404.

Cagetti, Marco, and Mariacristina De Nardi. 2008. "Wealth Inequality: Data and Models." *Macroeconomic Dynamics* 12 (Supplement 2): 285–313.

Calvet, Laurent, John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115 (5): 707–747.

Campbell, John Y. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review: Papers and Proceedings* 106 (5): 1–30.

Campbell, John Y. 2018. *Financial Decisions and Markets: A Course in Asset Pricing.* Princeton University Press, Princeton, NJ.

Campbell, John Y., Tarun Ramadorai, and Benjamin Ranish. 2014. "Getting Better or Feeling Better? How Equity Investors Respond to Investment Experience." National Bureau of Economic Research Working Paper 20000.

Campbell, John Y., Tarun Ramadorai, and Benjamin Ranish. 2018. Appendix to "Do the Rich Get Richer in the Stock Market? Evidence from India," available online at authors' websites.

Carhart, Mark M., 1997, "On Persistence in Mutual Fund Performance", *Journal of Finance* 52, 57–82.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. "Heterogeneity and Persistence in Returns on Wealth." National Bureau of Economic Research Working Paper 22822.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.

Kaniel, Ron, Gideon Saar, and Sheridan Titman. 2008. "Individual Investor Trading and Stock Returns." *Journal of Finance* 63: 273–310.

Markowitz, Harry M. 1976. "Investment for the Long Run: New Evidence for an Old Rule." Journal of Finance 31 (5): 1273–1286.

Martins, Renato, Harminder Singh, and Sukanto Bhattacharya. 2012. "What Does Volume Reveal: A Study of the Indian Single-Stock Futures Market." *Indian Journal of Economics & Business* 11: 409–419.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Cambridge, MA: Harvard University Press.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131 (2): 519–578.

Vashishtha, Ashutosh, and Satish Kumar. 2010. "Development of Financial Derivatives Market in India - A Case Study." 2010. *International Research Journal of Finance and Economics* 37: 15–29.

Figure 1A: Density of Log Account Value

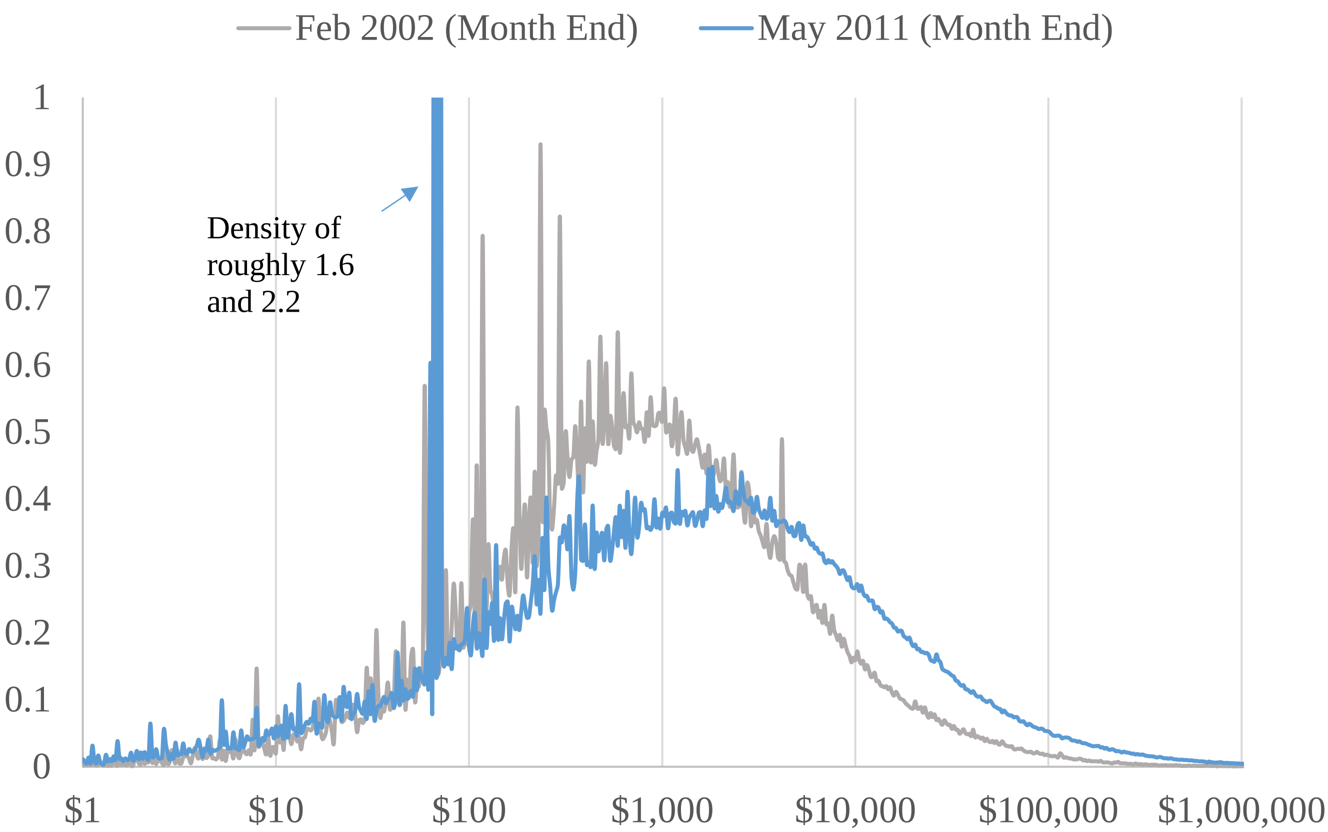

Figure 1B: Cumulative Distribution of Account Value

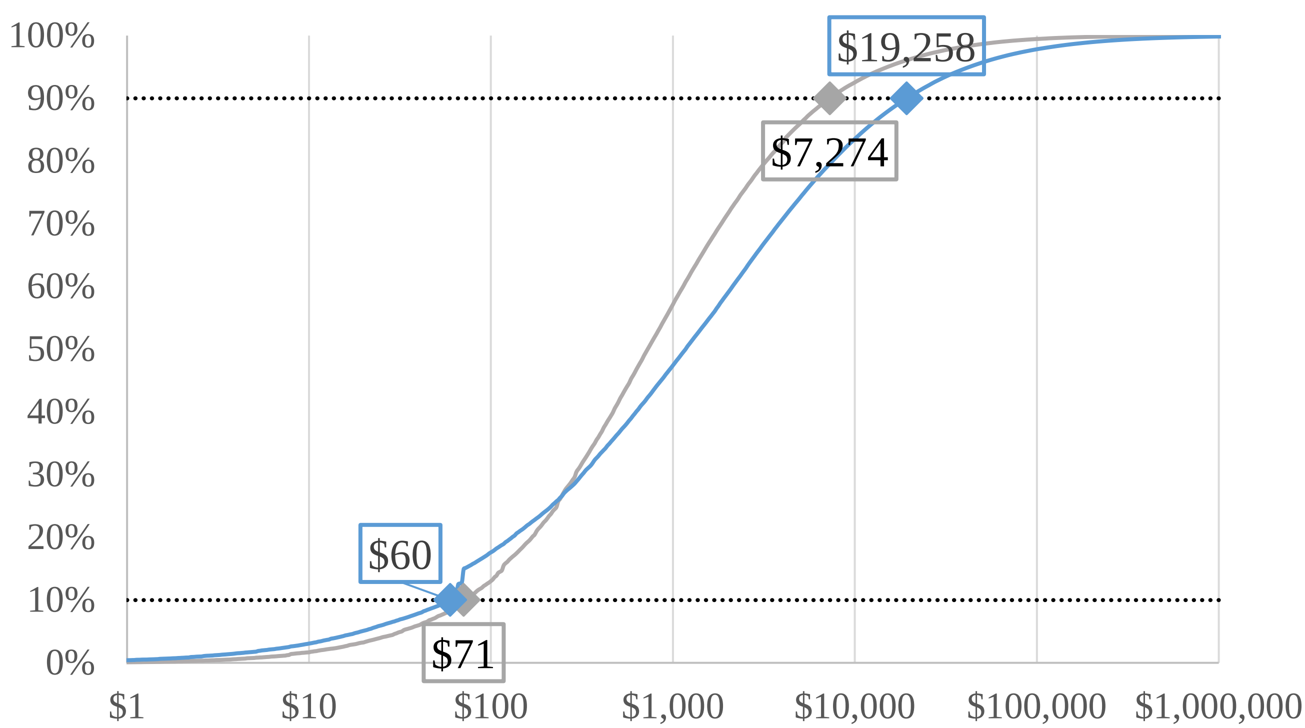

Table 1: Summary Statistics: March 2002 - May 2011

Panel A presents means, minimums, maximums, and standard deviations of cross-sectional statistics spanning March 2002 through May 2011. Panel B presents the time-series mean of cross-sectional correlations. Account growth, log return, and inflow statistics are from the subset of accounts for which all three variables are available.

Panel A: Time-series Variation in Cross-sectional Statistics

	Mean	Minimum	Maximum	Standard Deviation
Accounts in Sample	73,852	39,340	107,947	22,669
Account Entrance Rate	2.79%	0.79%	15.06%	2.31%
Account First-Time Entrance Rate	2.01%	0.43%	11.49%	1.81%
Account Exit Rate	1.85%	0.62%	4.79%	0.92%
Mean Log Account Value	10.80	10.32	11.53	0.26
Mean Account Growth	0.0167	-0.3331	0.3116	0.0850
Mean Log Account Returns	0.0138	-0.3708	0.3594	0.1015
Mean Log Net Inflows	0.0029	-0.0689	0.1355	0.0355
StDev Log Account Value	2.11	1.85	2.38	0.14
StDev Account Growth	0.4108	0.2753	0.6261	0.0794
StDev Log Account Returns	0.0913	0.0511	0.1584	0.0233
StDev Log Net Inflows	0.4054	0.2692	0.6263	0.0798

Panel B: Average Cross-sectional Correlations

	Lagged Log Account Value	Account Growth	Log Account Returns	Log Gross Inflows
Lagged Log Account Value	1.000	-0.088	0.027	-0.093
Account Growth		1.000	0.167	0.973
Log Account Returns			1.000	-0.058

Table 2: Return Factor Loadings by Account Value Deciles

Each column of this table presents statistics from the period March 2002 through May 2011. Panel A presents coefficients from regressions of monthly cross-sectional average excess returns on (lagged) account-value sorted accounts on four Fama French risk factors. Panel B shows the average number of stocks held for the account-value decile as well as the characteristics of the stockholdings of the decile. For the regressions in Panel A, account returns are based on the portfolio held at the end of the previous month and assume that the account does not trade during the month. Excess returns are constructed by subtracting the three-month Indian Treasury Bill rate. Risk factors are constructed from Indian equity data following the methodology described on Ken French's website. The stock characteristics in Panel B are time-series means of cross-sectional dollar-weighted median characteristic values. This measurement is robust to extreme outlier characteristics. Estimated beta comes from a regression of realized beta on two past years of realized beta as well as size, value, and momentum deciles and industry dummies. Standard errors are presented in parentheses throughout. Statistical significance of regression coefficients at the 5% and 10% level is indicated by bold and italicized type respectively.

Panel A: Risk Factor Loadings	Account Value Deciles										Largest minus Smallest
	Smallest	2	3	4	5	6	7	8	9	Largest	
Excess Return	**2.99%**	1.85%	**2.03%**	*1.90%*	*1.83%*	*1.80%*	**1.79%**	**1.76%**	**1.76%**	**1.70%**	-1.29%
	(1.35%)	(1.13%)	(1.03%)	(0.97%)	(0.94%)	(0.93%)	(0.91%)	(0.89%)	(0.86%)	(0.81%)	(0.78%)
Four-Factor Alpha	-0.17%	-0.38%	-0.01%	0.12%	0.18%	0.19%	0.20%	0.23%	0.27%	0.40%	0.57%
	(0.71%)	(0.67%)	(0.54%)	(0.51%)	(0.47%)	(0.45%)	(0.43%)	(0.41%)	(0.38%)	(0.36%)	(0.46%)
Market Beta	**1.21**	**1.15**	**1.13**	**1.10**	**1.09**	**1.08**	**1.08**	**1.06**	**1.04**	**1.00**	**-0.21**
	(0.05)	(0.08)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
Size (SMB)	**0.57**	**0.26**	**0.14**	**0.12**	**0.10**	**0.08**	**0.07**	**0.07**	**0.05**	**0.06**	**-0.51**
	(0.05)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
Value (HML)	**0.45**	0.17	0.16	0.08	0.04	0.03	0.02	0.00	-0.01	-0.07	**-0.51**
	(0.21)	(0.19)	(0.14)	(0.14)	(0.12)	(0.12)	(0.11)	(0.11)	(0.10)	(0.09)	(0.16)
Momentum (MOM)	**-0.36**	**-0.19**	**-0.21**	**-0.19**	**-0.18**	**-0.17**	**-0.15**	**-0.14**	**-0.11**	**-0.07**	**0.29**
	(0.08)	(0.09)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)

Panel B: Number of Stocks Held and Their Characteristics	Account Value Deciles									
	Smallest	2	3	4	5	6	7	8	9	Largest
Average Number of Stocks Held	1.56	1.89	2.71	3.57	4.84	6.38	8.58	11.58	16.21	28.89
	(0.01)	(0.03)	(0.02)	(0.02)	(0.04)	(0.06)	(0.08)	(0.14)	(0.20)	(0.30)
Estimated Beta	1.09	1.04	1.03	1.02	1.02	1.01	1.01	1.00	1.00	0.98
	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Market Capitalization (USD, millions)	114	3709	1149	1639	1766	2067	2378	2596	2961	3353
	(11)	(481)	(75)	(112)	(118)	(148)	(172)	(187)	(202)	(209)
Book to Market	0.97	0.71	0.69	0.66	0.65	0.63	0.61	0.59	0.57	0.52
	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Month t-12:t-2 Returns	11.8%	17.0%	24.4%	25.5%	26.5%	27.6%	28.5%	29.5%	30.6%	33.0%
	(3.7%)	(3.5%)	(3.8%)	(3.7%)	(3.7%)	(3.7%)	(3.6%)	(3.6%)	(3.5%)	(3.5%)

Table 3: Risk and Returns by Account Value Deciles

This table presents average monthly risk and return measures over the period March 2002 through May 2011 by account value deciles. These deciles are defined by the value of stock holdings at the end of the previous month. Account returns are constructed on the basis of these portfolios under the assumption that the account does not trade during the following month. Excess returns are constructed by subtracting the three-month Indian Treasury Bill rate, and are further adjusted across the three panels. Panel A reports in-sample realized excess returns. Panel B subtracts the sample-specific part of mean factor returns, taking long-run global factor returns from Ken French's website (over the period November 1990 through November 2017) and using the estimated risk factor loadings in Table 2. Panel C further subtracts the average part of the return associated with the estimated in-sample four-factor alpha reported in Table 2. Since these variations adjust only mean returns, excess return volatility is unaffected. Bootstrap standard errors are reported in parentheses, and reflect uncertainty about in-sample and global risk factor prices and in-sample factor loadings. Statistical significance at the 5% and 10% level is indicated by bold and italicized type respectively.

Risk and Returns by Size:

	Account Value Deciles										Largest minus Smallest
	Smallest	2	3	4	5	6	7	8	9	Largest	
Panel A: Realized:											
Excess Return Volatility	**23.7%**	**17.1%**	**15.8%**	**14.4%**	**13.6%**	**13.0%**	**12.5%**	**12.1%**	**11.6%**	**11.0%**	**-12.7%**
	(1.2%)	(1.0%)	(0.9%)	(0.9%)	(0.9%)	(0.9%)	(0.9%)	(0.9%)	(0.8%)	(0.8%)	(0.9%)
Excess Returns	**2.99%**	1.85%	**2.03%**	*1.90%*	*1.83%*	*1.80%*	**1.79%**	**1.76%**	**1.76%**	**1.70%**	-1.29%
	(1.35%)	(1.13%)	(1.03%)	(0.97%)	(0.94%)	(0.93%)	(0.91%)	(0.89%)	(0.86%)	(0.81%)	(0.78%)
Sharpe Ratio	**0.13**	*0.11*	**0.13**	**0.13**	**0.13**	**0.14**	**0.14**	**0.15**	**0.15**	**0.15**	0.03
	(0.05)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.04)
Excess Log Returns (X100)	0.72	0.52	0.87	0.91	0.94	0.99	1.02	1.05	1.10	1.10	0.38
	(1.26)	(1.13)	(1.01)	(0.96)	(0.94)	(0.92)	(0.91)	(0.89)	(0.86)	(0.81)	(0.69)
Panel B: Long Run Global Factor Prices:											
Excess Returns	0.42%	0.19%	0.52%	0.61%	0.66%	0.67%	0.68%	0.71%	*0.76%*	**0.86%**	0.44%
	(0.75%)	(0.70%)	(0.59%)	(0.56%)	(0.53%)	(0.52%)	(0.50%)	(0.48%)	(0.45%)	(0.44%)	(0.46%)
Sharpe Ratio	0.02	0.01	0.03	0.04	0.05	0.05	0.05	0.06	*0.07*	*0.08*	**0.06**
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)
Excess Log Returns (X100)	**-1.89**	-1.16	-0.66	-0.38	-0.24	-0.15	-0.08	0.00	0.10	0.26	**2.15**
	(0.76)	(0.73)	(0.61)	(0.57)	(0.54)	(0.53)	(0.51)	(0.49)	(0.46)	(0.45)	(0.46)
Panel C: Long Run Global Factor Prices, No Alpha:											
Excess Returns	*0.60%*	**0.57%**	*0.53%*	*0.49%*	*0.48%*	*0.48%*	*0.49%*	*0.48%*	*0.48%*	**0.47%**	-0.13%
	(0.32%)	(0.29%)	(0.27%)	(0.27%)	(0.26%)	(0.26%)	(0.26%)	(0.25%)	(0.25%)	(0.24%)	(0.14%)
Sharpe Ratio	*0.03*	**0.03**	*0.03*	*0.03*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Excess Log Returns (X100)	**-1.71**	**-0.78**	**-0.64**	*-0.50*	-0.42	-0.34	-0.28	-0.24	-0.18	-0.14	**1.57**
	(0.37)	(0.32)	(0.30)	(0.29)	(0.29)	(0.28)	(0.28)	(0.27)	(0.27)	(0.25)	(0.20)

Table 4: Decomposition of Inequality Growth (Change in the Variance of Log Account Value)

This table presents time-series averages of terms from cross-sectional variance decompositions. Panel A uses the identity $w_{i,t+1} = w_{it} + r_{i,t+1} + f_{i,t+1}$ to decompose the average monthly change in variance of log account value over the period March 2002 through May 2011 into components due to the variance of log account returns ($r_{i,t+1}$), net flows ($f_{i,t+1}$), and their covariances with each other and the previous log account value w_{it}. The remaining portion of the realized change in variance of log account value, $x_{i,t+1}$, is due to the entrance and exit of accounts. Panel B splits log returns into an expected and idiosyncratic component, $r_{i,t+1} = \mu_{it} + \varepsilon_{i,t+1}$, and uses these to further decompose the contribution of returns to the change in variance of log account value. Expected log returns are modeled in three ways. In model 1, expected log returns equal the mean realized log return of the given account value decile. Model 2 subtracts the part of returns due to in-sample factor prices from a Fama French four-factor model, while model 3 further subtracts the in-sample alpha (both as in Table 3). Panel C decomposes the contribution of net flows into the contribution of net flows variance and its covariance with the previous log account value.

Panel A: Overall Decomposition		Average Share of Change in Log Wealth Variance
Investment Returns: $Var_t^*(r_{i,t+1})+2Cov_t^*(w_{it},r_{i,t+1})$	0.0166	84.25%
Net Flows: $Var_t^*(f_{i,t+1})+2Cov_t^*(w_{it},f_{i,t+1})$	0.0079	40.27%
Covariance of Returns and Flows: $2Cov_t^*(r_{i,t+1},f_{i,t+1})$	-0.0046	-23.22%
Account Entry and Exit: $x_{i,t+1}$	-0.0003	-1.30%
Realized Change in Log Wealth Variance	0.0197	100.00%

Panel B: Investment Returns Component

	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Mean Realized Return	Long-Run Global Factor Prices	Long-Run Global Factor Prices, No Alpha	Mean Realized Return	Long-Run Global Factor Prices	Long-Run Global Factor Prices, No Alpha
$Var_t^*(\mu_{it})$	0.000004	0.000037	0.000018	0.02%	0.22%	0.11%
$Var_t^*(\varepsilon_{i,t+1})$	0.008870	0.008890	0.008879	53.54%	53.65%	53.59%
$2Cov_t^*(\mu_{it},\varepsilon_{i,t+1})$	0.000000	-0.000052	-0.000024	0.00%	-0.32%	-0.14%
$2Cov_t^*(w_{it},\mu_{it})$	0.006796	0.022407	0.014862	41.02%	135.24%	89.70%
$2Cov_t^*(w_{it},\varepsilon_{i,t+1})$	0.000899	-0.014713	-0.007168	5.42%	-88.80%	-43.26%
Total:	0.016568	0.016568	0.016568	100.00%	100.00%	100.00%

Panel C: Net Flows Component

$Var_t^*(f_{i,t+1})$	0.170739	2155.86%
$2Cov_t^*(w_{it},f_{i,t+1})$	-0.162820	-2055.86%
Total:	0.007920	100.00%