AER: Insights 2019, 1(2): 225–240
https://doi.org/10.1257/aeri.20180158

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

By John Y. Campbell, Tarun Ramadorai, and Benjamin Ranish*

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 equity 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. (*JEL* D14, D31, G11, O16)

New methods for imputing the wealth distribution have provided evidence that wealth inequality is increasing both in the United States and globally (Alvaredo et al. 2017; 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.[1] This paper uses detailed administrative data on the equity portfolios of Indian stock market investors to show that heterogeneous investment returns account for 84 percent of the increase in inequality of wealth held in equities between 2002 and 2011.[2]

India is a large country with a rapidly growing middle class that is starting to use risky financial markets for the first time. During our sample period, Indians primarily invest directly in stocks rather than through mutual funds or diversified

* Campbell: Department of Economics, Littauer Center, Harvard University, Cambridge, MA 02138, and NBER (email: john_campbell@harvard.edu); Ramadorai: Imperial College London, London SW7 2AZ, UK (email: t.ramadorai@imperial.ac.uk); Ranish: Federal Reserve Board, 20th and C Streets NW, Washington, DC 20551 (email: ben.ranish@frb.gov). Amy Finkelstein was the coeditor for this article. 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.

† Go to https://doi.org/10.1257/aeri.20180158 to visit the article page for additional materials and author disclosure statement(s).

[1] Piketty (2014, p. 430) 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."

[2] Although directly held equity wealth is only one component of total wealth, it is plausible that log equity wealth and log total wealth are positively correlated in India as elsewhere. Data from the All India Debt and Investment Survey, conducted as part of the Indian National Sample Survey in 2012, confirm a correlation of about 0.3 and an elasticity of total wealth with respect to equity account size of about 0.15 among stockholding households. We report details of this analysis in our online Appendix.

retirement savings vehicles (Badarinza, Balasubramaniam, and Ramadorai 2017). In these respects, India has much in common with other developing countries but differs from developed countries (Badarinza, Campbell, and Ramadorai 2016). We find that returns on directly held stocks generate slower growth of account value for small investors than for larger investors, because small Indian investors are poorly diversified.[3] This illustrates that low-cost diversification vehicles can benefit small investors and reduce the growth of inequality in emerging markets.

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.[4] 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 difference between the two is one-half the variance of the log return.[5] 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 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 two channels. First, undiversified risk-taking 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, average log returns can vary across investors who pursue different investment strategies. The second channel is particularly important if average log returns are correlated with initial wealth levels. Within this channel, cross-sectional variation in average log returns may reflect both variation in average simple returns—resulting from heterogeneity in investors' willingness to take risk, their ability to identify compensated risk exposures, or their stock-picking skill—and variation in the wedge between average simple returns and average log returns caused by portfolio variance.

[3] Comparable results are not available for the United States because of data limitations, but several papers have documented return heterogeneity in Norway (Fagereng et al. 2016) and Sweden (Calvet, Campbell, and Sodini 2007 and Bach, Calvet, and Sodini 2018a—henceforth, BCS). Because Swedish investors are relatively well diversified through mutual funds, 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.

[4] 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).

[5] 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) provides a textbook exposition.

We find that these two channels are roughly equally important in India. Heterogeneous log returns contribute to inequality in account size both through the random realizations of underdiversified portfolio returns and through higher average log returns that larger investors earn on their equity investments. Crucially, this is not because larger Indian investors have higher average simple returns. In fact, the opposite is true in our sample, because smaller investors have higher loadings on compensated risk factors including the market, small-stock, and value factors.[6] Rather, larger investors are better diversified so their idiosyncratic risk is lower, creating a smaller wedge between average simple and average log returns.

I. Data

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.

Our data on Indian equity accounts 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 percent market share of total assets tracked and a 60 percent 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 nonfinancial 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 percent. In addition, roughly 60 percent of mutual funds in India are held by corporations. Campbell, Ramadorai, and Ranish (2014)—henceforth, CRR—estimate that individuals' indirect equity holdings through mutual funds and related products were between 6 percent and 19 percent of total household equity holdings over the sample

[6] There is suggestive evidence that wealthier investors have stock-picking skills 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.

period. They also note that a 2009 SEBI survey found that about 65 percent 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 percent in 2002 to about 70 percent in 2011. The share of this held in individual accounts declines from about 20 percent to about 10 percent, 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 percent.[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 percent 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 percent of an individual account's stock holdings.[8] The online Appendix provides further details on data sources.

A. *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 percent per month, and the exit rate 1.9 percent per month, but the entry rate in particular is highly

[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 provided that at least some returns are available for the account. In about 2.5 percent 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 "entry and exit" component of our variance decomposition.

TABLE 1—SUMMARY STATISTICS: MARCH 2002–MAY 2011

	Mean	Minimum	Maximum	Standard deviation
Panel A. Time-series variation in cross-sectional statistics				
Accounts in sample	73,852	39,340	107,947	22,669
Account entry rate	2.79%	0.79%	15.06%	2.31%
Account first-time entry 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	0.017	−0.333	0.312	0.085
Mean log account returns	0.014	−0.371	0.359	0.101
Mean log net inflows	0.003	−0.069	0.136	0.035
Standard deviation log account value	2.11	1.85	2.38	0.14
Standard deviation account growth	0.411	0.275	0.626	0.079
Standard deviation log account returns	0.091	0.051	0.158	0.023
Standard deviation log net inflows	0.405	0.269	0.626	0.080
	Lagged log account value	Account growth	log account returns	log gross inflows
Panel B. Average cross-sectional correlations				
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

Notes: 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.

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 percent per month, very similar to the cross-sectional mean log return of 1.4 percent on investments in these accounts. The difference between these 2 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 1-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 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

[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" entry rate as 2.0 percent, implying that the average re-entry rate is 0.8 percent.

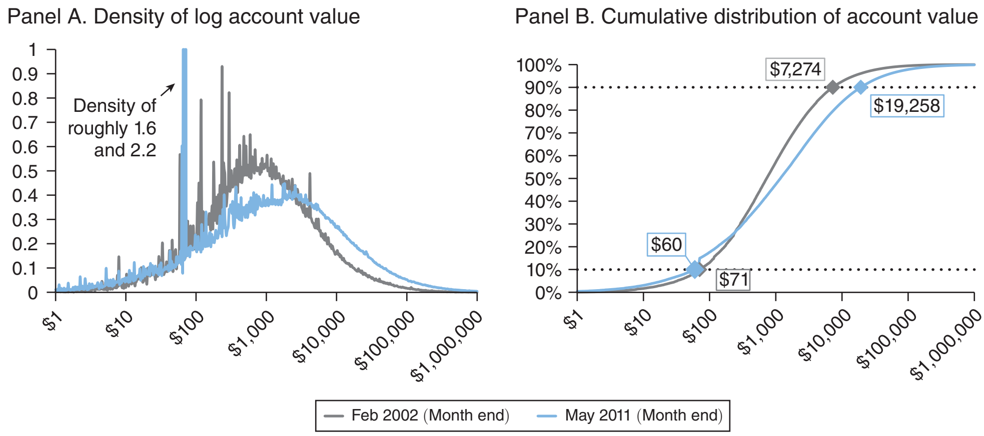

Figure 1

in standard amounts to many small accounts that hold undiversified single-IPO positions. Converting account sizes to US dollars, the tenth percentile of account size fell during our sample period from \$71 to \$60, while the ninetieth 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.

II. 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 percent per month in this sample period, while the largest earn only 1.70 percent per month. The difference in excess returns between the 2 extreme portfolios is −1.29 percent, but the estimate is noisy with a standard error of 0.78 percent. The large standard error reflects our

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

Table 2—Return Factor Loadings by Account Value Deciles

	Account value deciles										
	Smallest	2	3	4	5	6	7	8	9	Largest	Largest− smallest
Panel A. Risk factor loadings											
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											
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)	
Mkt cap (USD, MM)	114	3,709	1,149	1,639	1,766	2,067	2,378	2,596	2,961	3,353	
	(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%)	

Notes: 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 (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). 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.

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.

Table 2 uses a standard 4-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 2) for a sample of US retail investors. However, the effects of account size on these tilts differ from those reported by BCS (2018b) 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 4-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 percent per month, while the largest accounts have a positive alpha of 0.40 percent per month. These point estimates suggest that larger Indian investors have stock-picking skills relative to the four-factor Fama and French (1993) and Carhart (1997) model. However, the alpha spread of 0.57 percent has a large standard error of 0.46 percent, 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 volatility of account returns and the wedge between average simple and average log returns. We examine this in Table 3. The first row of Table 3 repeats the average excess simple return from the first row of Table 2. The second row of Table 3 reports the average excess log return. Unlike the average excess simple return, the average excess log return is increasing in account value, 0.72 percent per month for the smallest accounts and 1.10 percent per month for the largest accounts, although the difference of 0.38 percent has a standard error of 0.69 percent.

The difference between average simple and log returns is almost exactly equal to one-half the variance of log returns, as can be verified using the third row of Table 3 which reports the standard deviation of portfolio returns. Small Indian investors hold highly volatile portfolios, with an average standard deviation of 23.7 percent per month (equivalent to 82 percent per year). The portfolios of the largest

Table 3—Risk and Returns by Account Value Deciles

Risk and returns by size	Smallest	2	3	4	5	6	7	8	9	Largest	Largest−smallest
Panel A. Realized											
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%)
Excess log returns	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%)
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%)
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)
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%)
Excess log returns	−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%)
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)
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%)
Excess log returns	−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%)
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)

Notes: 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 (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). 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.

Indian investors have a much lower standard deviation of 11.0 percent per month (38 percent per year), and the difference of −12.7 percent has a standard error of only 0.9 percent. Thus, the better diversification of large investors more than offsets their low average simple returns, enabling large investors to earn higher average log returns than small investors.

The fourth row of Table 3 takes the ratio of the average excess simple return to standard deviation to calculate the Sharpe ratio. This is 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.

A. *Correcting for Luck*

One reason why small 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 percent 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 percent per month. Similarly, the average returns on SMB and HML were 0.91 percent per month and 2.49 percent per month in India during our sample, but only 0.06 percent and 0.33 percent per month in the longer run global data. The fourth factor, MOM, delivered 0.64 percent per month in India, which is barely above the longer run global average of 0.60 percent, 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 simple returns, average excess log returns, and Sharpe ratios 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, average excess simple returns are increasing in account value and in the bottom panel they are almost flat. In either case, average excess log returns and Sharpe ratios are increasing in account size. The difference in average excess log returns between the largest and smallest accounts is 2.15 percent in the first case, with a standard error of 0.46 percent, and 1.57 percent in the second case, with a standard error of 0.20 percent.

III. Decomposition of the Increase in Account Size 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 V_{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

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

where $V^0_{i,t+1} = V_{it}(1+R_{i,t+1})$ 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

$$(2)\qquad v_{i,t+1} = v_{it} + r_{i,t+1} + f_{i,t+1},$$

where

$$v_{it} = \log(V_{it}), \quad r_{i,t+1} = \log(1 + R_{i,t+1}), \quad \text{and} \quad f_{i,t+1} = \log(1 + F_{i,t+1}/V^0_{i,t+1}).$$

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

$$\begin{aligned} \text{(3)} \qquad \text{var}^*_t(v_{i,t+1}) - \text{var}^*_t(v_{it}) = {} & \text{var}^*_t(r_{i,t+1}) + 2\text{cov}^*_t(v_{it}, r_{i,t+1}) \\ & + \text{var}^*_t(f_{i,t+1}) + 2\text{cov}^*_t(v_{it}, f_{i,t+1}) \\ & + 2\text{cov}^*_t(r_{i,t+1}, f_{i,t+1}) + x_{i,t+1}. \end{aligned}$$

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 contributions of these terms to the change in the cross-sectional variance of log account size, which averages 0.0197 per month in our data. The contribution of log return inequality is 84 percent of this; the contribution of flow inequality is 40 percent; the interaction effect is -23 percent; and the effect of account entry and exit is a modest -1 percent of the total.[11] Thus, the dominant contributor to the increase in the inequality of account size in our data is indeed the heterogeneity in log investment returns.

Panel B looks separately at the two components of log return inequality (the first two terms on the right-hand side of (3)). The cross-sectional variance of log return and its cross-sectional covariance with log account value are roughly equally important, contributing 54 percent and 46 percent of the log return inequality effect.

Panels C and D look more closely at the covariance term. We have used accounts' market values to measure their size even though returns affect subsequent market values. One might be concerned that large accounts earn higher average log returns because there is momentum in underlying stock returns or persistent cross-sectional variation in investment skill or diversification, and so high log returns in the past explain both current account size and future average log returns. To address this concern, in panel C we break account size into a "book value" component that reflects only past contributions and proportional outflows, accumulated at a money market interest rate, and the residual which reflects past realized account returns. We find

[11] It may be surprising that the flow contribution is so modest when the cross-sectional volatility of flows is so large in Table 1. The explanation is 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.

TABLE 4—DECOMPOSITION OF INEQUALITY GROWTH (CHANGE IN THE VARIANCE OF LOG ACCOUNT VALUE)

				0.0197	
Realized change in log account value variance $\text{var}_t^*(v_{i,t+1}) - \text{var}_t^*(v_{it})$				Average share of change in log account value variance	
Panel A. Overall decomposition					
$\text{var}_t^*(r_{i,t+1}) + 2\text{cov}_t^*(v_{it}, r_{i,t+1})$	Returns	[I]		84.2%	
$\text{var}_t^*(f_{i,t+1}) + 2\text{cov}_t^*(v_{it}, f_{i,t+1})$	Net flows	[II]		40.3%	
$2\text{cov}_t^*(r_{i,t+1}, f_{i,t+1})$	Covariance of returns and flows	[III]		−23.2%	
$x_{i,t+1}$	Account entry and exit	[IV]		−1.3%	
Panel B. Returns component (panel A, item [I])				0.0166	
$\text{var}_t^*(r_{i,t+1})$		[A]		53.6%	
$2\text{cov}_t^*(v_{it}, r_{i,t+1})$		[B]		46.4%	
Panel C. Account value and returns covariance component (panel B, item [B])				0.0077	
$2\text{cov}_t^*(b_{it}, r_{i,t+1})$		[1]		69.8%	
$2\text{cov}_t^*(v_{it} - b_{it}, r_{i,t+1})$		[2]		30.2%	
Panel D. Account value and returns covariance component (panel B, item [B])				0.0077	
$2\text{cov}_t^*(v_{it}, \ln(\overline{R_{it}}))$		[a]		−107.6%	
$2\text{cov}_t^*(v_{it}, r_{i,t+1} - \ln(\overline{R_{it}}))$		[b]		207.6%	
Panel E. Returns component (panel A, item [I])				0.0166	
			Model 1	Model 2	Model 3
			Mean realized return	Long-run global factor prices	Long-run global factor prices, no alpha
$\text{Var}_t^*(\mu_{it})$		[i]	0.0%	0.2%	0.1%
$\text{Var}_t^*(\varepsilon_{i,t+1})$		[ii]	53.5%	53.7%	53.6%
$2\text{Cov}_t^*(\mu_{it}, \varepsilon_{i,t+1})$		[iii]	0.0%	−0.3%	−0.1%
$2\text{Cov}_t^*(v_{it}, \mu_{it})$		[iv]	41.0%	135.2%	89.7%
$2\text{Cov}_t^*(v_{it}, \varepsilon_{i,t+1})$		[v]	5.4%	−88.8%	−43.3%

Notes: This table presents time-series averages of terms from cross-sectional variance decompositions. Panel A uses the identity $v_{i,t+1} = v_{it} + r_{i,t+1} + f_{i,t+1}$ to decompose the average monthly change in variance of log account value over the period March 2002 through May 2011 into the share 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 v_{it}. The remaining portion of the realized change in variance of account value, $x_{i,t+1}$, is due to the entry and exit of accounts. Panel B decomposes the contribution of heterogeneous returns into components related to [A] undiversification in general and [B] the covariance of log account value and returns. Panel C decomposes the log account value and log returns covariance component into parts related to [1] the log book value of the account b_{it} (the account value under the counterfactual of three-month Indian treasury returns, and withdrawals set as the same proportion of account value as actual withdrawals) and [2] a remainder representing cumulative additional returns. Panel D alternatively decomposes the log account value and returns covariance component into parts related to [a] log mean (raw) returns and [b] the difference between log returns and log mean raw returns. Panel E 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 heterogeneous returns. 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).

that the book value covariance is 70 percent of the total, so the endogeneity of log account size—while not negligible—is a secondary effect. We report details of this calculation in the online Appendix.

In panel D we once again highlight the difference between average log return and average simple return. We show that the covariance between log account size and average simple return is negative, but outweighed by the positive covariance between log account size and the wedge between average log return and average simple return. This result confirms the patterns reported in Table 3.

A. *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

$$\begin{aligned} \text{var}_t^*(r_{i,t+1}) + 2\text{cov}_t^*(v_{it}, r_{i,t+1}) = {} & \text{var}_t^*(\mu_{it}) + \text{var}_t^*(\varepsilon_{i,t+1}) \\ & + 2\text{cov}_t^*(\mu_{it}, \varepsilon_{i,t+1}) \\ & + 2\text{cov}_t^*(v_{it}, \mu_{it}) + 2\text{cov}_t^*(v_{it}, \varepsilon_{i,t+1}). \end{aligned} \tag{5}$$

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 account size and expected log return; this captures the tendency of larger/richer accounts to invest more effectively and earn higher log returns. The fifth term is the covariance between log account size and unexpected log return; this can be nonzero in any cross section, if investment strategies favored by wealthy accounts do better or worse than average, but should average to zero in a long sample period if μ_{it} is a rational expectation.

Panel E 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, while the cross-sectional variance of unexpected log returns and the covariance between log account size and realized log return contribute roughly equally as shown earlier in panel B of Table 4.

Where the models differ is in the breakdown of the covariance term into the covariance of log account size 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). Thus 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 equity 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 account size and expected log return is even larger and the covariance between log account size 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 equity 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 inequality by 36 percent to 75 percent of the total observed increase in inequality in this period.

IV. Robustness

In the online Appendix, we show that our results are robust to several variations in our empirical methodology. Other measures of account size inequality also trend upward in our sample period. Focusing on the right tail of the size distribution, we examine deciles within the top 5 percent 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 upward by survivorship bias or bid-ask bounce (Blume and Stambaugh 1983), and exclude accounts that hold predominantly IPO stocks, and again find similar results.

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 equity 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 percent contribution of account entry and exit reported in panel A of Table 4.

V. 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 percent 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 the inequality of account size 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 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.

Our paper 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. Our results also highlight the importance for developing countries of investment vehicles, such as mutual funds and exchange traded funds, that are already common in developed countries and that give small investors an affordable way to diversify risk.

REFERENCES

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2017. *World Inequality Report 2018*. World Inequality Lab.

Bach, Laurent, Laurent Calvet, and Paolo Sodini. 2018a. "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2706207 (accessed May 9, 2019).

Bach, Laurent, Laurent Calvet, and Paolo Sodini. 2018b. "Internet Appendix for Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." https://sites.google.com/site/laurentbach/AppendixRichPickings.pdf (accessed May 9, 2019).

Badarinza, Cristian, Vimal Balasubramaniam, and Tarun Ramadorai. 2017. "The Indian Household Finance Landscape." *India Policy Forum 2016-17*. New Delhi: National Council of Applied Economic Research.

Badarinza, Cristian, John Y. Campbell, and Tarun Ramadorai. 2016. "International Comparative Household Finance." *Annual Review of Economics* 8 (1): 111–44.

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 (2): 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 (3): 387–404.

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

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini. 2007. "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes." *Journal of Political Economy* 115 (5): 707–47.

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

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

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. 2019. "Do the Rich Get Richer in the Stock Market? Evidence from India: Dataset." *AER: Insights*. https://doi.org/10.1257/aeri.20180158.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1): 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 (1): 273–310.

Markowitz, Harry M. 1976. "Investment for the Long Run: New Evidence for an Old Rule." *Journal of Finance* 31 (5): 1273–86.

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 (2): 409–19.

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): 519–78.

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.