

# Turnover, account value and diversification of real traders: evidence of collective portfolio optimizing behavior

David Morton de Lachapelle<sup>1,2,4</sup> and Damien Challet<sup>3</sup>

<sup>1</sup> Department of Quantitative Asset Management, Swissquote Bank SA, Geneva Office, 1196 Gland, Switzerland

<sup>2</sup> LBS—Institute of Theoretical Physics, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

<sup>3</sup> Physics Department, Fribourg University, 1700 Fribourg, Switzerland

E-mail: [david.mortondelachapelle@swissquote.ch](mailto:david.mortondelachapelle@swissquote.ch), [david.morton@epfl.ch](mailto:david.morton@epfl.ch) and [damien.challet@unifr.ch](mailto:damien.challet@unifr.ch)

**Abstract.** Despite the availability of very detailed data on financial markets, agent-based modeling is hindered by the lack of information about real trader behavior. This makes it impossible to validate agent-based models, which are thus reverse-engineering attempts. This work is a contribution towards building a set of stylized facts about the traders themselves. Using the client database of Swissquote Bank SA, the largest online Swiss broker, we find empirical relationships between turnover, account values and the number of assets in which a trader is invested. A theory based on simple mean-variance portfolio optimization that crucially includes variable transaction costs is able to reproduce faithfully the observed behaviors. We finally argue that our results bring to light the collective ability of a population to construct a mean-variance portfolio that takes into account the structure of transaction costs.

<sup>4</sup> Author to whom any correspondence should be addressed.

## Contents

<b>1. Introduction</b>	<b>2</b>
<b>2. Description of the data</b>	<b>3</b>
<b>3. Results</b>	<b>4</b>
3.1. Account values . . . . .	4
3.2. Mean turnover . . . . .	5
3.3. Mean turnover versus account value . . . . .	7
3.4. Turnover rescaled by account value . . . . .	9
<b>4. The influence of transaction costs on trading behavior: optimal mean-variance portfolios</b>	<b>11</b>
4.1. Nonlinear relationship between account value and number of assets . . . . .	14
<b>5. Turnover, number of assets and account value</b>	<b>16</b>
<b>6. Discussion and outlook</b>	<b>20</b>
<b>Acknowledgment</b>	<b>20</b>
<b>References</b>	<b>20</b>

### 1. Introduction

The availability of large data sets on financial markets is one of the main reasons behind the number and variety of works devoted to their analysis in various fields, and especially so in econophysics, since physicists very much prefer to deal with very large data sets. At the macroscopic level, the analysis of millions of tick-by-tick data points uncovered striking regularities of price, volume, volatility and order booking dynamics (see [1]–[4] for reviews). Since these phenomena are caused by the behavior of individual traders, news and the interplay between the two, finding a microscopic mechanism that allows agent-based models to reproduce some of these stylized facts is an important endeavor meant to give us insight into the causes of large fluctuations, be it herding [5], competition for predictability [6], portfolio optimization leading to market instability [7] or chaotic transitions [8].

Market phenomenology appears as a typical example of collective phenomena to the eyes of statistical physicists. Thus, the temptation to regard the numerous power laws found in empirical works as signatures of criticality is intense. But if the former are really due to a phase transition, one wishes at least to know what the phases are, which is hard to guess from the data alone. According to early herding theoretical models [5], the phase transition may lie in the density of social communication and imitation, and is of percolation type, thereby linking power-law distributed price and volume, criticality and agent behavior. The standard Minority Game (MG) [9] also has a single phase transition point where market predictability is entirely removed by the agents, without any specular effect on price and volume. On the other hand, grand-canonical MGs [10]–[13] that allow the agents not to play have a semi-line of critical points that do produce stylized facts of price, volume and volatility dynamics. In the framework of statistical physics, the phase transition is due to symmetry breaking, i.e. it is a transition between predictable and perfectly efficient markets. This also suggests that the emergence of large fluctuations is due to market efficiency.

There are, of course, many other possible origins of power laws in financial markets that have nothing to do with a second-order phase transition. The simplest mechanism is to consider multiplicative random walks with a reflecting boundary [14]. Long-range memory of volatility is well reproduced in agent-based models whose agents act or do nothing depending on a criterion based on a random walk [15]. Assuming pre-existing power-law distributed wealth, an effective theory of market phenomenology links the distributions of price returns, volume and trader wealth [16]. On the other hand, markets are able to produce power-law distributed price returns by simple mechanisms of limit order placement and removal without the need for wealth inequality [17, 18]. However, in turn, one needs to explain why limit orders are placed in such a manner. The heterogeneity of time scales may provide an explanation for order placement far away from best prices if power law distributed [19], but additional work is needed in order to explain order placement near best prices, which causes these large price moves. Finally, a recent simple model of investment with leverage is able to reproduce some stylized facts [20].

But mechanisms alone may not be sufficient to replicate the full complexity of financial markets, as some part of it may lie instead in the heterogeneity of the agents themselves. While the need for heterogeneous agents in this context is intuitive (see, e.g., [21]), there are no easily available data against which to test or to validate microscopically an agent-based model. Even if it is relatively easy to design agent-based models that reproduce some of the stylized facts of financial markets (see, e.g., [6, 8], [22]–[24]), one never knows whether this is achieved for good reasons, except for volatility clustering [15]: it is to be expected that real traders behave sometimes at odds with one’s intuition. Thus, without data about the traders themselves, one is left with the often frustrating and time-consuming task of reverse-engineering the market in order to determine the good ingredients indirectly. Some progress has been made recently with the analysis of transactions in the Spanish stock market aggregated by brokers [25], hence with mesoscale resolution.

Data on trader behavior are found in the files of brokers, usually shrouded in secrecy. But this lack of data accessibility is not entirely to blame for the current ignorance of real-trader dynamics: researchers, even when given access to broker data, have focused on trading gains and behavioral biases, often with factor-based analyses (see, e.g., [26]–[28]).

We aim at providing a coherent picture of how various types of traders behave and interact, making it possible for agent-based models to rest on a much more solid basis. This paper is the first of a series that will establish stylized facts about trader characteristics and behavior. One of the most important aspects of these papers will be to characterize the heterogeneity of the traders in all respects (account value, turnover, trading frequency, behavioral biases, etc) and the relationships between these quantities in probability distribution, not with factors. This paper is first devoted to the description of the large data set that we use. It then focuses on the relationship between trader account value, turnover per transaction and transaction costs, both empirically and theoretically. We will show that, while the traders have a spontaneous tendency to build equally weighted portfolios, the number of stocks in a portfolio increases nonlinearly with their account value, which we link to portfolio optimization and broker transaction fee structure.

## 2. Description of the data

Our data are extracted from the database of the largest Swiss online broker, Swissquote Bank SA (further referred to as Swissquote). The sample contains comprehensive details about all the

19 million electronic orders sent by 120 000 professional and non-professional online traders from January 2003 to March 2009. Of these orders, 65% have been canceled or have expired and 30% have been filled; the remaining 5% were still valid as of 31 March 2009. Since this study focuses on turnover as a function of account value, we chose to exclude orders for products that allow traders to invest more than their account value, also called leveraging, i.e. orders to margin-calls markets, such as the foreign exchange market (FOREX) and the derivative exchange EUREX. The resulting sample contains 50% of orders for derivatives, 40% for stocks and 4% for bonds and funds. Finally, 70% of these orders were sent to the Swiss market, 20% to the German market and about 10% to the US market.

Swissquote clients consist of three main groups: *individuals*, *companies* and *asset managers*. Individual traders, also referred to as *retail clients*, are mainly non-professional traders acting on their own account. The accounts of companies are usually managed by individuals trading on behalf of a company and, as we shall see, they behave very much like retail clients, albeit with a larger typical account value. Finally, asset managers manage accounts of individuals and/or companies, some of them dealing with more than a thousand clients. Their behavior differs markedly from that of the other two categories of clients.

### 3. Results

#### 3.1. Account values

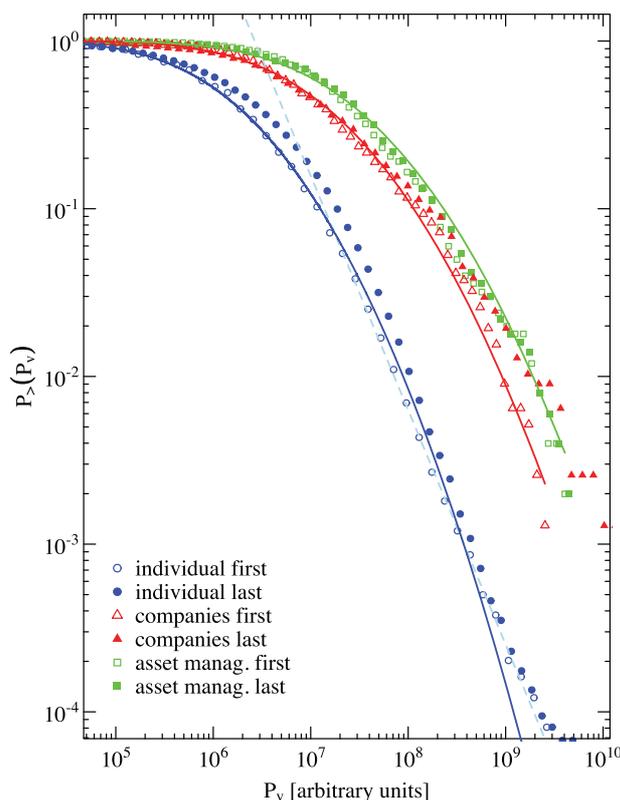
Numerous studies have been devoted to the analysis and modeling of wealth dynamics and distribution among a population (see [30] and references therein). The general picture is that, in a population, a very large majority lies in the exponential part of the reciprocal cumulative distribution function (RCDF), while the wealth of the richest people is Pareto-distributed, i.e. according to a power law.

The account value of Swissquote traders is by definition the sum of all their assets (cash, stock, bonds, derivatives, funds, deposits) and is denoted  $P_v$ . In order to simplify our analysis, we compute  $P_v$  once per day after US markets close and take this value as a proxy for the next day's account value. Figure 1 displays this distribution computed at the time of the first and last transactions of the clients. Results are shown for the three main categories of clients. Maximum likelihood fits to the tail of the individual traders to the Pareto model  $p(x) \sim (x/x_{\min})^{-\gamma}$  were performed using the  $BC_a$  bootstrap method of [31] and determining the parameter  $x_{\min}$  by minimizing the Kolmogorov–Smirnov statistics as in [29]. Results are reported in table 1.

The values of  $\gamma$  are in line with the wealth distribution of all major capitalistic countries (see [32] for a possible origin of Pareto exponents between 2.3 and 2.5). Thus the retail clients are most probably representative of the Swiss population. The account value distributions of companies and asset managers have no clear power-law tails, in agreement with the results of a recent model that suggests a log-normal distribution of mutual fund asset sizes [33, 34]. Consequently, figure 1 also reports a fit of the data to log-normal distributions  $\ln N(\mu, \sigma^2)$ , which approximate more faithfully  $P_>(P_v)$  than the Student and the Weibull distributions for the three categories of clients, except its extreme tail, in the case of retail clients.

**Table 1.** Results of the fits of the Pareto law  $(x/x_{\min})^\gamma$  to the account value  $P_v$  of individuals.

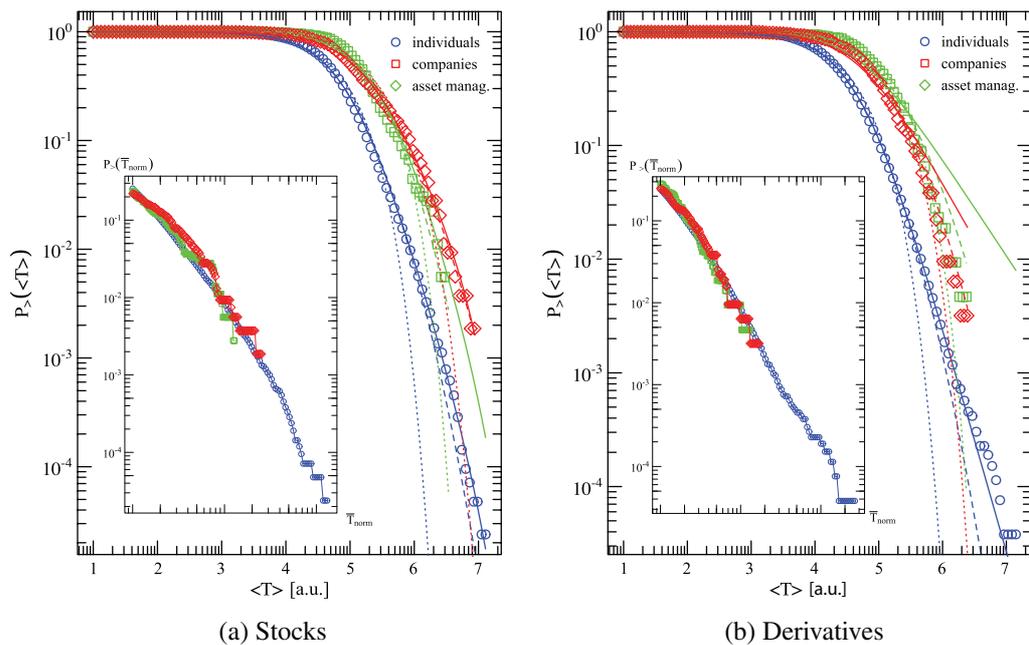
Individuals	$\gamma$	$x_{\min}$
First transaction	$2.33 \in [2.29, 2.37]_{95}$	$2.30 \times 10^6 \in [1.99 \times 10^6, 2.59 \times 10^6]$
Last transaction	$2.39 \in [2.33, 2.44]_{95}$	$3.73 \times 10^6 \in [3.15 \times 10^6, 4.29 \times 10^6]$



**Figure 1.** RCDF of the portfolio value  $P_v$  for the three categories of clients at the time of their first (empty symbols) and last (filled symbols) transactions. Several models have been fitted to the data by maximum likelihood estimation (MLE): the Student distribution (Pareto with plateau), the Weibull (stretched exponential) and the log-normal distribution. The best candidate, determined graphically and via bootstrapping the Kolmogorov–Smirnov test [29], was found to be the log-normal distribution, which is the only one shown here for the sake of clarity. The dashed line in light blue results from an MLE fit to the tail of the individual traders with the Pareto distribution  $p(x) \sim (x/x_{\min})^{-\gamma}$  (see section 3.1).

### 3.2. Mean turnover

The turnover of a single transaction  $i$ , denoted by  $T_i$ , is defined as the price paid times the volume of the transaction and does not include transaction fees. We have excluded the traders that have leveraged positions on stocks, hence  $T_i \leq P_v$ . More generally one wishes to determine how the average turnover of a given trader relates to his portfolio value. In passing, since  $P(P_v)$  has fat tails, the only way the distribution of  $T$  can avoid having fat tails is if the typical turnover



**Figure 2.** Reverse cumulative distribution function of the mean turnover per transaction for the three categories of clients and for both stock and derivative transactions. In the insets, the tail part of the RCDF of  $\langle T_{\text{norm}} \rangle = \langle T \rangle / \text{mean}(\langle T \rangle)$ . The solid curves are maximum likelihood fits to (1) for stocks and (2) for derivatives. The dotted lines are fits to the Weibull distribution and the dashed lines to the log-normal distribution (table 3).

**Table 2.** Parameter values and 95% confidence intervals for the MLE fit of the account values to the log-normal distribution  $\ln N(\mu, \sigma^2)$ . For each category of investors, the first and second rows correspond to the account value at the time of the first and last transactions, respectively (see text). Note that portfolio values have been multiplied by an arbitrary number for confidentiality reasons. This only affects the value of  $\mu$ .

	$\mu$	$\sigma$
Individuals	$13.94 \pm 0.02$	$2.87 \pm 0.01$
	$14.25 \pm 0.02$	$2.01 \pm 0.01$
Companies	$16.0 \pm 0.2$	$2.0 \pm 0.1$
	$15.9 \pm 0.2$	$2.4 \pm 0.1$
Asset managers	$16.7 \pm 0.2$	$1.8 \pm 0.1$
	$16.7 \pm 0.2$	$2.0 \pm 0.1$

is proportional to  $\log(P_v)$ . We denote by  $\langle T \rangle$  the mean turnover per transaction for a given client over the history of his activities.

Figure 2 reports its RCDF for stocks and derivatives for the three categories of clients; all RCDFs have a first plateau and then a fat tail. For stocks, the tails are not pure power laws, but they are for derivatives. Indeed, fitting the RCDFs with Weibull, log-normal and

**Table 3.** Results of the maximum likelihood fit of  $P_{>}(\langle T \rangle)$  with (1) and (2) for the three categories of clients. The 95% confidence intervals reported in smaller character are computed by the biased-corrected accelerated ( $BC_a$ ) bootstrap method of [31].

	Stocks (1)		Derivatives (2)
	$\gamma$	$\beta \times 10^{-6}$	$\gamma$
Individuals	1.97 [1.83, 2.10]	0.98 [0.46, 1.5]	1.98 [1.91, 2.15]
Companies	1.29 [1.52, 1.89]	1.66 [0.44, 2.3]	–
Asset managers	1.93 [1.47, 2.93]	0.91 [–7.8, 4.5]	–

Zipf–Mandelbrot distributions with an exponential cut-off, defined as

$$F_{>}^{(1)}(x) = \frac{c^\gamma e^{-\beta x}}{(c+x)^\gamma}, \tag{1}$$

clearly shows that the latter is the only one that does not systematically underestimate the tail of the RCDF for stocks. Estimated values of  $\beta$  and  $\gamma$  are given in table 3.

The RCDFs related to the turnover of transactions on derivative products have clearer power-law tails for retail clients, which we fitted with a standard Zipf–Mandelbrot function, defined as

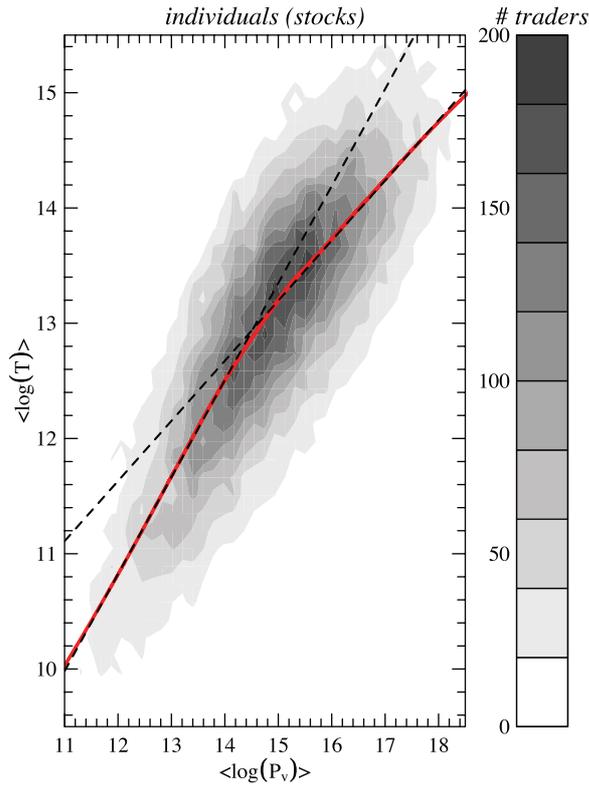
$$F_{>}^{(2)}(x) = \frac{c^\gamma}{(c+x)^\gamma}. \tag{2}$$

The parameters estimated are to be found in table 3. Because of the power-law nature of this tail, fits with Weibull and log-normal distributions are not very good in the tails. While the decision process that allocates a budget to each type of product may be essentially the same, the buying power is larger for derivative products, which may explain the absence of a cut-off. Fits for companies and asset managers is very difficult and mostly non-conclusive because of insufficient sample size. The good quality of the tail collapse (see inset) tends to indicate that the three distributions are identical, but we could not fit the RCDF of companies and asset managers with (2). As reported in figure 2(b), log-normal distributions are adequate choices in these cases. Since the quality of the fits is poor, we do not report the resulting parameters.

### 3.3. Mean turnover versus account value

The relationship between  $\langle T \rangle$  and  $\langle P_v \rangle$  is important as it dictates what fraction of their investable wealth the traders exchange in markets. We first produce a scatter plot of  $\langle \log T \rangle$  versus  $\langle \log P_v \rangle$  (figure 4). In a log–log scale plot, it shows a cloud of points that is roughly increasing. A density plot is, however, clearer for retail clients as there are many more points (figure 3).

These plots make it clear that there are simple relationships between  $\log T$  and  $\log P_v$ . A robust non-parametric regression method [34] reveals a double linear relationship between



**Figure 3.** Density plot of the average  $\log T$  versus the average  $\log P_v$ , robust non-parametric fit (red line) and linear fits (dashed lines).

$\langle \log T \rangle$  and  $\langle \log P_v \rangle$  for all three categories of investors (see figures 3 and 4),

$$\langle \log T \rangle = \beta_x \langle \log P_v \rangle + a_x, \quad (3)$$

where  $x = 1$  when  $\langle \log P_v \rangle < \Theta_1$  and  $x = 2$  when  $\langle \log P_v \rangle > \Theta_2$ . Fitted values with confidence intervals are reported in table 4.

This result is remarkable in two respects: (i) the double linear relation, not obvious to the naked eye, separates investors into two groups, and (ii) the ranges of values where the transition occurs is very similar across the three categories of traders.

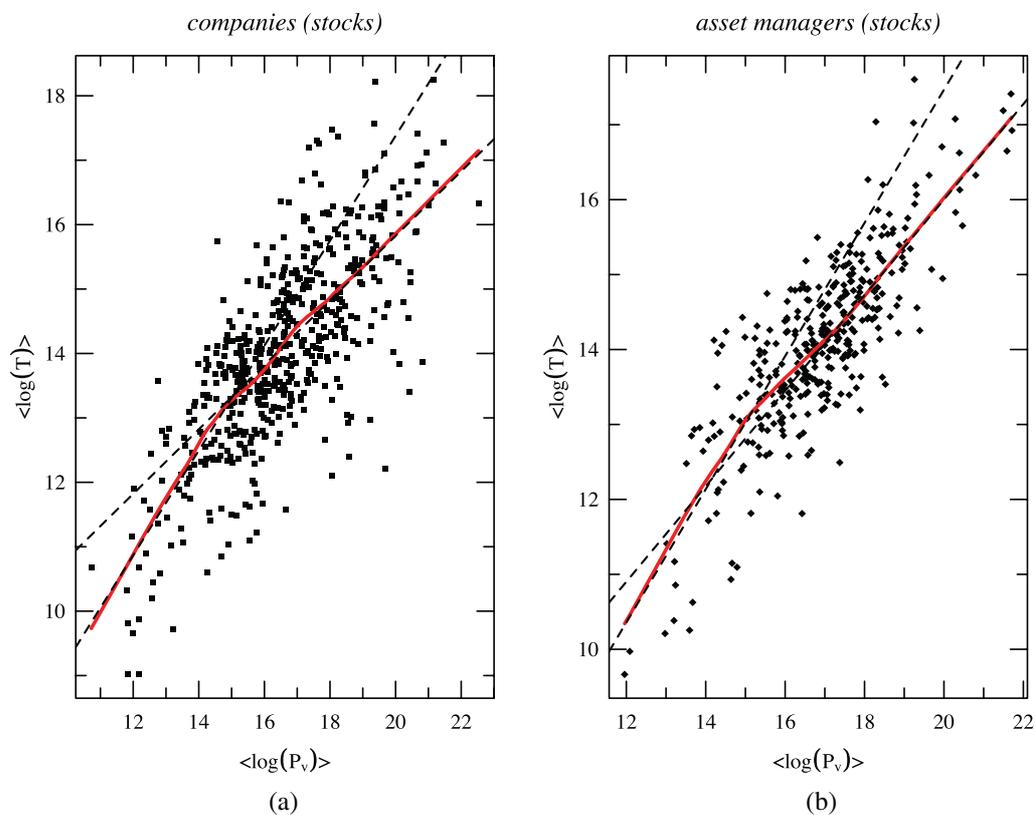
The relationships above only apply to averages over all the agents. This means that there are some intrinsic quantities that make all the agents deviate from this average line. Detailed examination of the regression residuals show that the latter are, for the most part (i.e. more than 95%), normally distributed with constant standard deviations  $\xi_x$  and that the residuals deviating from the normal distributions are not fat-tailed. This directly suggests the simple relation for individual traders

$$T^i = e^{a_x + \delta^i a_x} (P_v^i)^{\beta_x} \leq e^{\Theta_x}, \quad (4)$$

where  $T^i$  and  $P_v^i$  are, respectively, the turnover and portfolio value of investor  $i$ , and  $\delta^i a_x$  are i.i.d.  $N(0, \xi_x^2)$  idiosyncratic variations independent of  $P_v$  that mirror the heterogeneity of the

**Table 4.** Parameter values and 95% confidence intervals for the double linear model (4). For each category of investors, the first and second rows correspond, respectively, to  $\langle \log P_v \rangle \leq \Theta_1$  and  $\langle \log P_v \rangle \geq \Theta_2$ . For confidentiality reasons, we have multiplied  $P_v$  and  $T$  by a random number. This only affects the true values of  $a_x$  and  $\Theta$  in the table.

	$\beta_x$	$a_x$	$\xi$	$\Theta$	$R^2$
Individuals	$0.84 \pm 0.02$	$0.73 \pm 1.25$	0.71	14	0.52
	$0.54 \pm 0.01$	$5.07 \pm 0.15$	0.77	14.5	0.40
Companies	$0.81 \pm 0.13$	$1.12 \pm 8.17$	0.88	15.5	0.47
	$0.50 \pm 0.07$	$5.82 \pm 1.65$	1.00	15.6	0.33
Asset managers	$0.89 \pm 0.20$	$-0.31 \pm 0.76$	0.62	15.5	0.52
	$0.63 \pm 0.08$	$3.28 \pm 5.78$	0.62	16.5	0.46

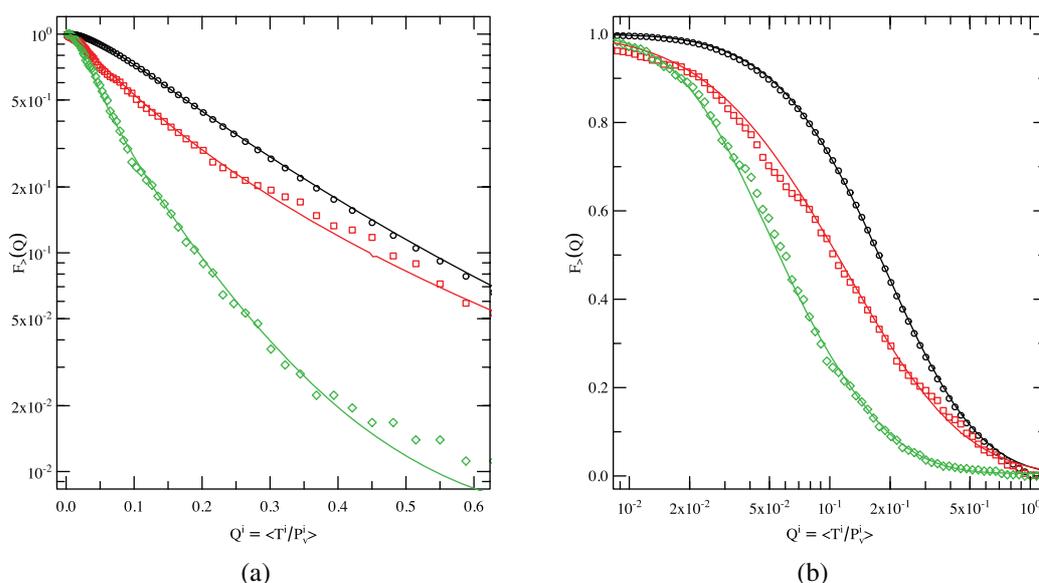


**Figure 4.** Density plot of the average  $\log T$  versus the average  $\log P_v$ , robust non-parametric fit (red line) and linear fits (dashed lines).

agents. As we shall see, portfolio optimization with heterogeneous parameters yields this precise relationship.

### 3.4. Turnover rescaled by account value

Let us now measure the typical fraction of wealth exchanged in a single transaction, defined as  $Q = \langle \frac{T}{P_v} \rangle$ . Since the inverse of this ratio is an indirect (and imperfect) proxy of the number  $N$



**Figure 5.** Reverse cumulative distribution function of  $Q = \langle \frac{T}{P_v} \rangle$ , the mean ratio of the turnover over the portfolio value for individual traders (black), companies (red) and asset managers (green). The left plot is in lin-log scale and the right plot is in log-lin scale. The solid lines come from theoretical predictions of section 3.4.2.

of assets that a trader owns, it also indicates how well diversified his investments are. Hence, it can be viewed as a simple proxy of the risk profiles of the agents.

**3.4.1. Data.** Figure 5 shows that the distributions look exponential to the naked eye for about 90% of the individuals and nearly 80% of the companies, while that of the asset managers is rapidly more complex than a simple exponential. We derive exact relationships for this quantity in section 3.4.2 that show that these distributions are in fact not exponential but log-normal.

The resulting picture is that only a small fraction of customers trade a large fraction of their wealth, on average. Interestingly, these figures show a clear difference between the three categories of clients. As discussed above, figure 5 roughly reflects the risk profile of the different types of customers: less than 10% of asset managers trade, on average, more than 20% of their clients' capital in a single transaction; this rises to 30% for companies and 45% for retail clients. Note, however, that, despite the fact that the account values of companies and asset managers are comparable, companies tend to have a  $Q$  closer to that of the individuals. This suggests either that companies hold a smaller  $N$  than asset managers for the same account value, or that asset managers tend to make smaller adjustments to the quantities of assets.

**3.4.2. Theory.** Since we know the distributions of  $T$ ,  $P_v$  and their relationship, we are in a position to derive analytical expressions for  $Q^i = \langle \frac{T(t)}{P_v(t)} \rangle$  of investor  $i$ . The distribution of  $Q$  across the population of online investors can be easily found using (4) and the distribution of  $P_v$ . Let  $P_{T,P_v}(t, p_v)$  denote the joint distribution of  $T$  and  $P_v$ :

$$P_{Q=T/P_v}(q) = \int_0^\infty p_v P_{T,P_v}(qp_v, p_v) dp_v = \int_0^\infty p_v P_{T|P_v}(qp_v|p_v) P_{P_v}(p_v) dp_v. \quad (5)$$

Let us now assume for the sake of clarity that  $T = e^{a+\delta a} P_v^\beta$ . Given  $P_v$ , the turnover  $T$  follows a log-normal distribution with mean  $\log p_v + a$  and variance  $\xi^2$ . Substituting  $P_{T|P_v}(t|p_v) = \ln N(\log p_v + a, \xi^2)$  in (5) leads, after some simplifications, to

$$P_Q(q) = \int_0^\infty \frac{1}{\sqrt{2\pi\xi^2q}} \exp\left(-\frac{(\log(qp_v^{1-\beta}) - a)^2}{2\xi^2}\right) P_{P_v}(p_v) dp_v, \quad (6)$$

and

$$F_Q(q) = \int_0^q P_Q(x) dx = \int_0^\infty \frac{1}{2} \operatorname{erfc}\left(\frac{a - \log(qp_v^{1-\beta})}{\sqrt{2}\xi}\right) P_{P_v}(p_v) dp_v, \quad (7)$$

where  $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2} dy$  is the complementary error function. As expected, when  $\beta = 0$  (i.e.  $T$  and  $P_v$  are independent), we recover the product of the two marginal distributions. On the other hand, when  $\beta = 1$ , i.e. when  $T$  is proportional to  $P_v$ ,  $P_Q(q) = \ln N(a, \xi^2)$ , which is the distribution of the factor  $e^{a+\delta a}$ . For other values of  $\beta$ , the functions  $P_Q$  and  $F_Q$  cannot be determined analytically unless  $P_{P_v}$  takes a particular form, as shown below. However, the moments of  $P_Q(q)$  can be arranged in a simpler form,

$$E(q^n) = \int_0^\infty q^n P_Q(q) dq = e^{na+(1/2)n^2\xi^2} \int_0^\infty \frac{1}{p_v^{n(1-\beta)}} P_{P_v}(p_v) dp_v, \quad (8)$$

that is, the (log-normal) moments of  $T/P_v$  times an integral term smaller or equal to 1 (because in practice  $P_{P_v}(p_v) > 1$ ).<sup>5</sup> Hence, the relation  $E(q^n) \leq e^{na+(1/2)n^2\xi^2}$  with equality when  $\beta = 1$  holds for any distribution of the account value  $P_v$ .

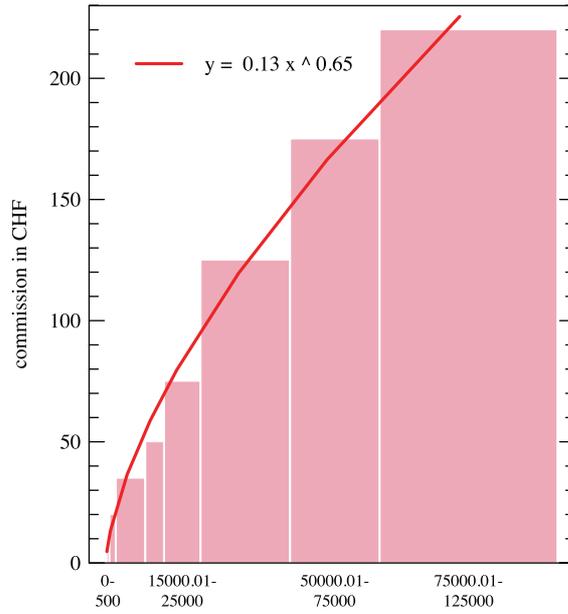
In section 3.1, we have shown that the distribution of  $P_v$  is well approximated by a log-normal distribution. This particular choice of distribution makes the previous integrals analytically tractable. Indeed, with  $P_{P_v} = \ln N(\mu, \sigma^2)$ , straight integration of (6) leads to  $P_Q = \ln N(M, S^2)$ , where  $M = a - (1 - \beta)\mu$  and  $S^2 = \xi^2 + (1 - \beta)^2\sigma^2$ . This simple result has some practical interest: given the distribution parameters and the coupling factor  $\beta$ , one can draw realistic  $q$  factors for agent-based modeling as  $Q = e^{M+SX}$ , where  $X$  is  $\mathcal{N}(0, 1)$  distributed. Furthermore, in the next section, we show how the value of  $\beta$  may be inferred from the transaction cost structure, which decreases the number of parameters to four.

Figure 5 confirms the validity of the above theoretical results, once expanded to the case of a bilinear relation between  $T$  and  $P_v$ . It is noteworthy that the continuous lines are no fits on empirical  $q$  factors, but use instead the results of the separate fits on the turnover and account distributions.

#### 4. The influence of transaction costs on trading behavior: optimal mean-variance portfolios

Apart from risk profiles, education and typical wealth, the differences in the turnover as a function of wealth observed above between the three populations of traders may also lie in the difference of their actual transaction cost structure. Swissquote current standard structure for the Swiss market (its shape is very similar for European and US markets) is shown in figure 6. It is a piece-wise constant, nonlinear-looking function. Fitting all segments to equation (10)

<sup>5</sup> Mathematically, all the moments of  $Q$  always exist since  $\beta \leq 1$  and  $P_v(p_v)$  must decay faster than  $p_v^{-1}$  to be a valid distribution.



**Figure 6.** Swissquote fee curve for the Swiss stock market. Commissions based on a sliding scale of costs are common practice in the world of online finance. The red line results from a nonlinear fit to equation (10). Parameter values are  $C = 0.13 \in [0.05, 0.5]_{95}$  and  $\delta = 0.63 \in [0.5, 0.74]_{95}$ , where the 95% confidence intervals are obtained from the  $BC_a$  bootstrap method of [31].

gives  $\delta = 0.63 \in [0.5, 0.74]_{95}$ . The fee structure of most brokers is not set in stone and can be negotiated. A frequent request is to have a flat fee, i.e. a fixed cost per transaction corresponding to a constant function. Since quite clearly the negotiation power of large clients or of clients that carry out many transactions is more important, asset managers are more likely to obtain a more favorable fee structure than basic retail clients.

Since buying some shares of an asset is the result of an unconscious or calculated portfolio construction process, one first needs a theoretical reference point with which to compare the population characteristics as measured in the previous subsection. In other words, we shall use results from portfolio optimization theory with nonlinear transaction cost functions to understand the results of the previous subsection.

Quite curiously, all analytical papers in the literature on optimal portfolios either neglect transaction costs or assume constant or linear transaction cost structures; nonlinear structures are tackled numerically. Thus, we incorporate the specific nonlinear transaction cost structure faced by the traders under investigation in the classic one-shot portfolio optimization problem studied by Brennan [35], who restricted its discussion to fees proportional to the number of securities, in other words, a flat fee per transaction.

Building optimal mean-variance stock portfolios consists, for a given agent, of selecting which stock to invest in and in what proportion by maximizing the expected portfolio growth, usually called return, while trying to reduce the resulting *a priori* risk. One cost function that corresponds to such requirements is

$$L_\lambda(R) = \lambda E(R) - \text{Var}(R), \quad (9)$$

where  $R$  is the stochastic return of the portfolio over the investment horizon (e.g. 1 month, 1 year) and  $\lambda$  tunes the trade-off between risk and return. As such, it can be interpreted as a measure of an investor's attitude towards risk: the larger  $\lambda$ , the more risk-averse the investor.

The return of the portfolio can be decomposed into contributions from risky assets (stocks, derivatives, etc), the interests of the amount kept in cash and the total relative cost of broker commission, which we denote as  $R = R^{\text{risky}} + R^{\text{cash}} - R^{\text{cost}}$ . Mathematically,

- $R^{\text{risky}} = \sum_{i=1}^N x_i R_i$ , where  $R_i$  is the return of stock  $i$  over this horizon,  $x_i$  is the fraction of the total wealth invested in this stock and  $N$  is the total number of investable assets; we shall denote the total fraction of wealth invested in risky assets by  $x = \sum_{i=1}^N x_i$ ;
- $R^{\text{cash}} = (1 - x)r$ , where  $r$  is the interest rate;
- $R^{\text{cost}} = \frac{\sum_{i=1}^N F(x_i P_v)}{P_v} (1 + r)$ , where  $F(x)$  is the amount charged by a broker to exchange an amount  $x$  of cash into shares, or vice versa.

The focus of this section is to derive explicit relationships between  $F$ , the number of assets to hold in a portfolio, and the account value  $P_v$ . Whereas previous works only considered special cases for  $F$  that are not compatible with the fees structure of Swissquote, we need to introduce a cost function that can accommodate all the standard broker commission schemes. The two extreme cases are (i) flat fee per transaction, i.e. a fixed cost that does not depend on the amount exchanged, and (ii) a proportional scheme, possibly with a maximum fee. Swissquote's standard scheme stands in between, and is well approximated by, a power law with a maximum fee  $F_{\text{max}}$ . We hence choose

$$F(x_i P_v) = \min(C(x_i P_v)^\delta, F_{\text{max}}), \quad (10)$$

where  $\delta$  interpolates between a flat fee ( $\delta = 0$ ), as in [35], and a proportional scheme ( $\delta = 1$ ) via a power law, and  $C$  is a constant.

Following the well-known one-factor model of Sharpe [36], we assume that the return of asset  $i$  follows the global market's return  $R_M$  with an idiosyncratic proportionality factor  $\beta_i$ . More specifically,

$$R_i = \beta_i (R_M - r) + r + \varepsilon_i, \quad (11)$$

where  $\varepsilon_i$  is an uncorrelated white noise  $E(\varepsilon_i) = E(\varepsilon_i \varepsilon_j) = E(R_M \varepsilon_i) = 0$ . This equation means that the systematic idiosyncratic part of  $R_i$  only applies to the return above the risk-free interest rate, also called the market risk premium.

This completely specifies the functional  $L_\lambda$ . Returning to (9), one first computes the expectation and variance of the portfolio return:

$$\begin{aligned} E(R) &= \sum_{i=1}^N x_i E(R_i) + (1 - x)r - \frac{\sum_{i=1}^N F(x_i P_v)}{P_v} (1 + r), \\ &= (E(R_M) - r) \sum_{i=1}^N x_i \beta_i + r - \frac{(1 + r)C}{P_v^{1-\delta}} \sum_{i=1}^N x_i^\delta, \end{aligned} \quad (12)$$

and

$$\begin{aligned} \text{Var}(R) &= \text{Var}(R^{\text{risky}}) \\ &= \text{Var}(R_M) \sum_{i=1}^N (x_i \beta_i)^2 + \sum_{i=1}^N x_i^2 \text{Var}(\varepsilon_i). \end{aligned} \quad (13)$$

Note that, since here the risk-free rate is non-random, the portfolio variance is independent of both the risk-free investment and broker commission. This does not hold for the expected return.

In principle, the functional  $L$  depends on  $N$ , the number of assets in the portfolio,  $\lambda$  the risk parameter and  $x_i$  the fraction of account value to invest in risky product  $i$ . Assuming that  $x_i$  is constant for all  $i$  (i.e. equally weighted allocation), we are left with only three parameters since  $x_i = x/N$ . Thus, from the optimization of the resulting functional, one can obtain a relationship between any two of these parameters. We are mostly interested in  $N$  as a function of  $x$ .

#### 4.1. Nonlinear relationship between account value and number of assets

We will first assume that agents seek the optimal fraction of their account value  $x^*$  to invest in  $N$  securities— $N$  being known—given the risk-free rate  $r$  and broker commission  $F(x_i W)$ . The optimal solution is simply obtained by setting  $x_i = x/N$  in (12) and (13) and by equating to zero the derivative of (9) with respect to  $x$ . This leads to the following transcendental equation for  $x^*$ ,

$$x^* = \frac{\lambda \bar{\beta}(E(R_M) - r) - \delta(1+r)C(N/x^* P_v)^{1-\delta}}{2 \left( \bar{\beta}^2 \text{Var}(R_M) + (1/N) \overline{\text{Var}}(\varepsilon) \right)}, \quad (14)$$

where  $\bar{\beta} = \frac{1}{N} \sum_{i=1}^N \beta_i$  and  $\overline{\text{Var}}(\varepsilon) = \frac{1}{N} \sum_{i=1}^N \text{Var}(\varepsilon_i)$  is the mean idiosyncratic volatility. Provided that the investor risk tolerance  $\lambda$  has been reliably estimated, which is usually a complex task [37], and that the Sharpe model is adequate, (14) can be used directly in a real-world portfolio optimization problem. The  $\beta_i$  and  $\varepsilon_i$  are then obtained by regressing the returns of all the stocks with (11). The optimal solution is expected to be reliable in the absence of significant residual correlations between  $\varepsilon_i$  and  $\varepsilon_j$ . In the more common situation where  $\lambda$  is unknown, one can derive a second equation for the optimal number of securities on the assumption that portfolios are sufficiently homogeneous, or that the investment horizon is long enough so as to have  $\bar{\beta}$  and  $\overline{\text{Var}}(\varepsilon)$  independent of  $N$ . As shown in figure 7,  $\bar{\beta}$  on the US stock market is persistently close to the one for various time horizons and values of  $N$ , consistently with the homogeneous assumption. Taking a few technical precautions into account [35], the differentiation of the Lagrangian (9) with respect to  $N$  leads to

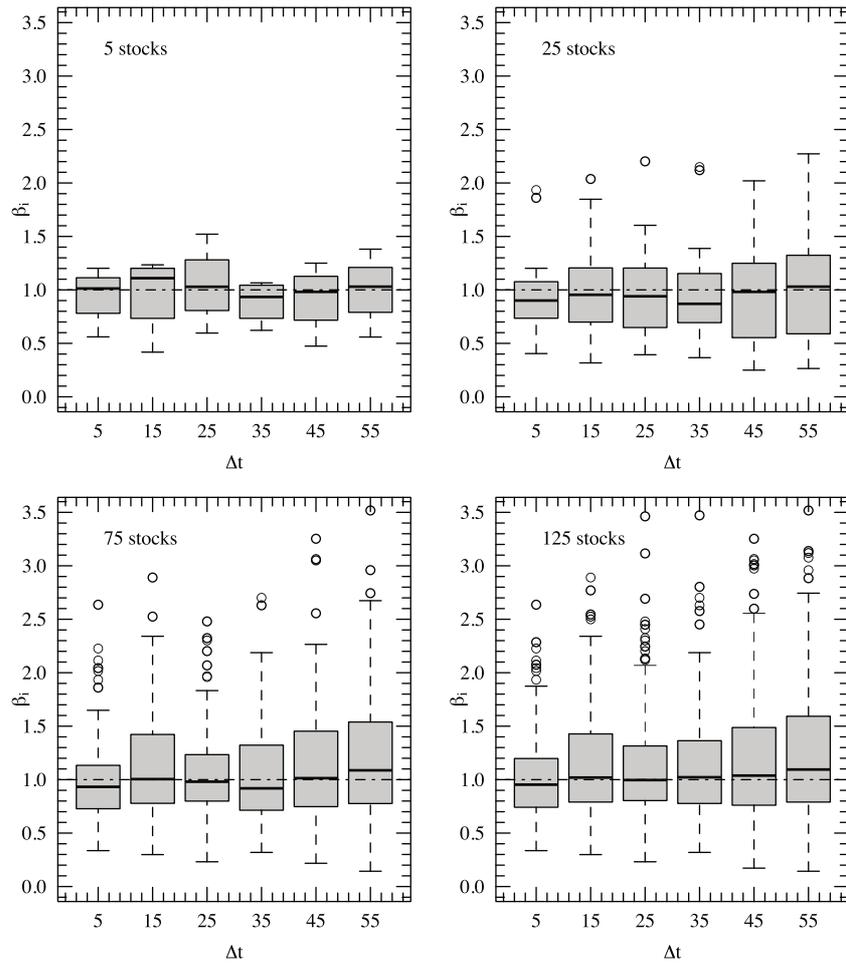
$$\lambda = \frac{\overline{\text{Var}}(\varepsilon) P_v^{1-\delta}}{(1-\delta)C(1+r) (N^*/x)^{2-\delta}}, \quad (15)$$

where it is assumed that  $\delta < 1$ , since for  $\delta = 1$  the optimum investment does not depend on  $N$  through the cost function. According to (15), the agent risk tolerance increases with their account value  $P_v$ , in agreement with various survey studies on the risk tolerance of actual investors (see the literature review [38]). Using (14) and (15) to get rid of  $\lambda$ , we obtain

$$N^{2-\delta} \left( 1 + \frac{\delta}{1-\delta} \frac{K}{N} \right) = K \frac{\bar{\beta}(E(R_M) - r)}{(1-\delta)C(1+r)} (x P_v)^{1-\delta}, \quad (16)$$

where  $K$  is the ratio of residual risk to market risk defined as

$$K = 2 \left( \frac{\bar{\beta}^2 \text{Var}(R_M)}{\overline{\text{Var}}(\varepsilon)} + \frac{1}{N} \right)^{-1} \underset{N \gg 1}{\approx} 2 \frac{\overline{\text{Var}}(\varepsilon)}{\bar{\beta}^2 \text{Var}(R_M)}. \quad (17)$$



**Figure 7.** Box plot of empirical  $\beta$ s obtained from the regression of several US stocks on the S&P500. The observation period covers 2001–2008 and returns are computed on various time horizons  $\Delta t$  (in days). Results show that  $\bar{\beta} = \frac{1}{N} \sum_{i=1}^N \beta_i \approx 1$  for all values of  $\Delta t$  and (even small)  $N$ , consistently with the homogeneous assumption of section 4.1.

Given the desired level of systematic risk  $x$ , (16) can be solved for  $N$  numerically in an actual portfolio optimization. Further insight is gained by considering the high diversification limit  $N \gg 1$ , which yields  $1 + \frac{\delta}{1-\delta} \frac{K}{N} \approx 1$  in (16) and thus

$$N = \left( K \frac{\bar{\beta}(E(R_M) - r)}{(1-\delta)C(1+r)} \right)^{1/(2-\delta)} (xP_v)^{(1-\delta)/(2-\delta)}, \quad (18)$$

where  $K$  is given by the right-hand side of (17). The latter equation generalizes [35] to the case of a varying cost impact represented here by the parameter  $\delta$  (i.e. the result of [35] is recovered by setting  $\delta = 0$  and  $\beta_i = 1$  in (18)). These results can be further generalized to non-equally weighted portfolios by differentiating (9) with respect to  $x_i$  and assuming again a homogeneous condition for the  $\beta_i$ s.

In essence, equation (18) says that the number of securities held in an equally weighted mean-variance portfolio with Sharpe-like returns is related to the amount invested as

$$\log(N) = \frac{1 - \delta}{2 - \delta} \log(x P_v) + \kappa \quad (19)$$

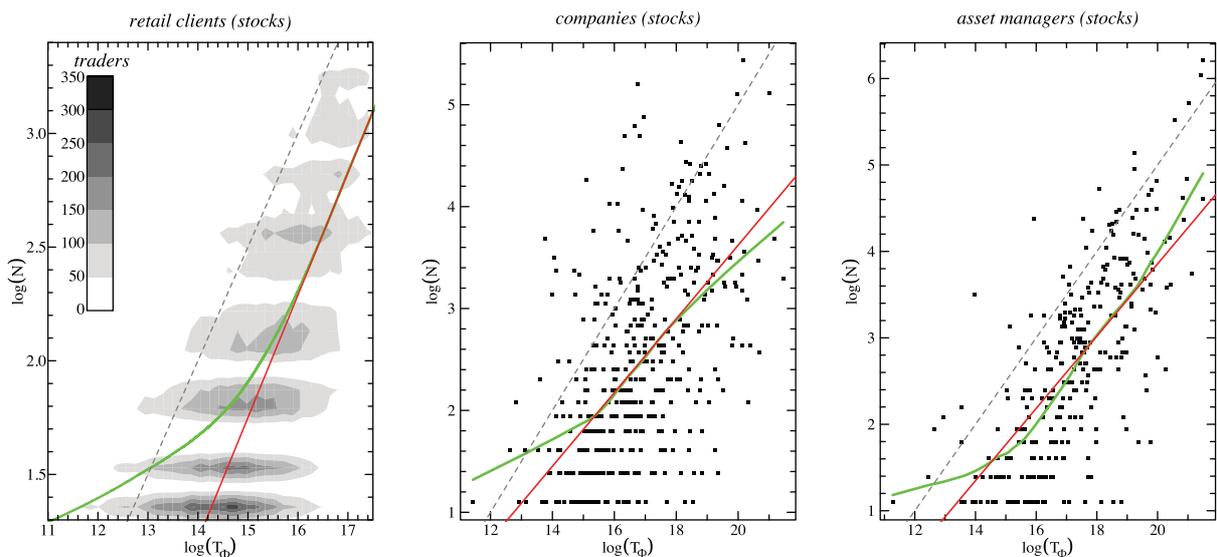
in the high diversification limit, where  $\kappa$  is the pre-factor of  $(x P_v)^{(1-\delta)/(2-\delta)}$  in (18). The last equation gives  $N$  as a function of  $P_v$  for a predefined  $x$  in the optimal portfolio. The heterogeneity of the traders, beyond their account value, is not apparent yet, but may occur both in  $x$  and  $\kappa$ . First each trader may have his own preference regarding the fraction of this account to invest in risky assets,  $x$ . Therefore one should replace  $x$  by  $x^i$ . Next,  $\kappa$  includes both a term related to transaction costs, which does vary from trader to trader, and some measures and expectation of market returns and variance. Each trader may have his own perception or way of measuring them. Hence  $\kappa$  should also be replaced by  $\kappa^i$ . Finally, both terms can be merged in the same constant term  $\zeta^i = \frac{1-\delta}{2-\delta} \log(x^i) + \kappa^i$ . This explains how the heterogeneity of the traders is the cause of fluctuations in the kind of relationships we are interested in.

## 5. Turnover, number of assets and account value

The above result only links  $N$  with  $P_v$ , but one also wishes to obtain relationships that involve the turnover per transaction,  $T$ . Whereas in section 3 we have characterized the turnover of any transaction, the results of section 4 rest on the assumption that the agents build their portfolio by selecting a group of assets and stick to them over a period of time. This, obviously, does not include the possibility of speculating by a series of buy and sell trades on even a single asset, nor portfolio rebalancing, which consists of adjusting the relative proportions of some assets. We thus have to find a way to differentiate between portfolio building, rebalancing and speculation. Here, we shall focus on portfolio building in order to test and link the results of section 4 to those of section 3.

We have found a simple, effective method that can separate portfolio-building transactions from the other ones. We assume that the transactions of trader  $i$  that correspond to the building of his portfolio are restricted to the first transaction of assets not traded previously. Sell orders are ignored, since Swissquote clients cannot short sell easily. In other words, if trader  $i$  owns some shares of assets A, B and C and then buys some shares of asset D, the corresponding transaction is deemed to contribute to his portfolio building process. The set of such transactions is denoted by  $\Phi_i$ , while the full set of transactions is denoted by  $\Omega_i$ . Any subsequent transaction of shares of assets A, B, C or D are left out of  $\Phi_i$ . The number of different assets that trader  $i$  owns is supposed to be  $N_i \simeq |\Phi_i|$ , where  $|X|$  is the cardinal of set  $X$ . This approach assumes that a trader always owns shares in all the assets ever traded. Surprisingly, this is by far the most common case. We shall drop the index  $i$  from now on.

Let us now focus on  $T_\Phi = \sum_{k \in \Phi} T_k$ , the total turnover that helped in building his portfolio. We should first check how it is related to the total portfolio value  $P_v$ . Let us define  $\langle P_v \rangle_\Phi$ , the account value of a trader averaged at the times at which he trades a new asset. Plotting  $\log \langle P_v \rangle_\Phi$  against  $\log T_\Phi$  gives a cloudy relationship, as usual, but fitting it with  $\log \langle P_v \rangle_\Phi = \chi \log T_\Phi$  gives  $\chi = 1.03 \pm 0.02$  for individuals,  $\chi = 0.99 \pm 0.02$  for asset managers and  $\chi = 1.00 \pm 0.01$  for companies with an adjusted  $R^2 = 0.99$  in all cases. This relationship trivially holds for the traders who buy all their assets at once, as assumed in the portfolio model. The traders who do not lie on this line either hold positions in cash (in which case this line is a lower bound) or



**Figure 8.** Turnover of transactions contributing towards building a portfolio  $T_\Phi$  versus the number  $N$  of assets held by a given trader at the time of the transaction. Green lines: non-parametric fit; red lines: fits of the linear part of the non-parametric fit. From left to right: companies, asset managers and individuals.

**Table 5.** Slope  $\alpha$  linking  $\log T_\Phi$  and  $\log N$  for the three trader categories.

	Individuals	Companies	Asset managers
$\alpha$	$0.52 \pm 0.02$	$0.36 \pm 0.14$	$0.44 \pm 0.13$
$\log T_\Phi \in$	[16, 19]	[17, 19.8]	[15.8, 18]

do not build their portfolio in a single day: they pile up positions in derivative products or stocks whose price fluctuations are the origin of the deviations from the line. But the fact that the slope is close to 1 means that the average fluctuation is zero and, hence, that, on average, trades do not make money from the positions taken on new stocks. The consequence of this is that  $\log P_v$  can be replaced by  $\log T_\Phi$  in (19), thus setting  $x = 1$ ,

$$\log N = \frac{1 - \delta}{2 - \delta} \log T_\Phi + \kappa. \quad (20)$$

The  $x = 1$  assumption is in fact quite reasonable: most Swissquote traders do not use their trading account as savings accounts and are fully invested. We do not know what amount they keep on their other bank accounts.

A robust non-parametric fit does reveal a linear relationship between  $\log N$  and  $\log T_\Phi$  in a given region  $(N, T_\Phi) \in \Gamma$  (figure 8, table 5). In this region, we have

$$\log N = \alpha \log T_\Phi + \beta, \quad (21)$$

which gives

$$\alpha = \frac{1 - \delta}{2 - \delta}. \quad (22)$$

**Table 6.** The results of the double linear regression of  $\log\langle T \rangle_\Phi$  versus  $\log\langle P_v \rangle_\Phi$ . For each category of investors, the first and second rows correspond, respectively, to  $\log\langle P_v \rangle_\Phi \leq \Theta_1$  and  $\log\langle P_v \rangle_\Phi \geq \Theta_2$ , where  $\Theta_{1,2}$  have been determined graphically using the non-parametric method of [34], as in section 3.3. Parameters are as in the double linear model (4). For confidentiality reasons, we have multiplied  $P_v$  and  $T$  by a random number, which only affects the true values of  $\Theta_{1,2}$  and of the ordinate  $a_x$ .

	$\beta_x$	$a_x$	$\xi$	$\Theta$	$R^2$
Individuals	$0.85 \pm 0.02$	$0.71 \pm 0.16$	0.65	14.5	0.59
	$0.51 \pm 0.01$	$5.62 \pm 0.17$	0.76	15	0.31
Companies	$0.83 \pm 0.17$	$1.03 \pm 2.47$	0.86	15.5	0.42
	$0.62 \pm 0.14$	$3.99 \pm 2.55$	0.93	17	0.32
Asset managers	$0.84 \pm 0.25$	$0.45 \pm 3.77$	0.79	15.95	0.50
	$0.73 \pm 0.17$	$1.72 \pm 3.23$	0.72	18	0.41

We still need to link  $\langle T \rangle_\Phi$  and  $\langle P_v \rangle_\Phi^\beta$ . While section 3 showed that the unconditional averages lead to  $\langle T \rangle \sim \langle P_v \rangle^\beta$ , one also finds that  $\langle T \rangle_\Phi \sim \langle P_v \rangle_\Phi^\beta$ . Therefore, one can write

$$\log \langle T \rangle_\Phi = \beta \log \langle P_v \rangle_\Phi + \text{const}, \quad (23)$$

where the fitted values are reported in table 6.

Thus, one is finally rewarded with the missing link

$$\beta = \frac{1}{2 - \delta}, \quad (24)$$

which directly involves the transaction cost structure in the relationship between turnover and portfolio value, as argued in section 3.<sup>6</sup> This relationship allows us to close the loop as we are now able to relate directly the exponents linking  $T$ ,  $N$  and  $P_v$ . Going back to section 3, one understands that the existence of a bilinear relationship between log-turnover and log-account values, i.e. of two values of  $\beta$  for each of the three categories of clients, is linked to two values of  $\delta$ : a flat fee structure or the disregard for transaction costs leads to  $\beta = \frac{1}{2}$ , while proportional fees ( $\delta = 1$ ) give  $\beta = 1$ .

Let us finally discuss the empirical values of  $\alpha$ ,  $\beta$  and  $\delta$  against their theoretical counterparts, which is summarized in table 7.

1. *Small values of  $T_\Phi$* : It was impossible to measure  $\alpha$  in that case since the non-parametric fit shows a nonlinear relationship in the log-log plot for retail clients, which we trust more since they have many more points than the graphs for the two other categories of clients. But it may not make sense to expect a linear relationship since such a relationship is only expected for  $N$  large enough ( $N \geq 10$  in practice) and a small  $T_\Phi$  is related to a small  $N$ . Thus, we can only test  $\beta = 1/(2 - \delta)$ . The reported value of  $\beta$  is consistent across all the clients. Retail clients have a larger  $\delta_{\text{eff}} = 2 - \frac{1}{\beta}$  than the estimated  $\delta_{\text{SQ}}$ . Since the shape of the fee structure is discontinuous, the values of these exponents can hardly be expected

<sup>6</sup> Note that this relationship can be obtained directly by assuming that all the transactions happen at the same time and, hence, that  $T = (x P_v)/N$ , which leads straightforwardly to (24).

**Table 7.** Table summarizing the empirical and theoretical relationships between  $\alpha$ ,  $\beta$  and  $\delta$ .

	Individuals	Companies	Asset managers
Small $T_\phi$			
$\beta$	$0.85 \pm 0.02$	$0.83 \pm 0.17$	$0.84 \pm 0.25$
$\log T_\phi <$	14.5	17	18
$\delta_{\text{eff}} = 2 - \frac{1}{\beta}$	$0.82 \pm 0.02$	$0.80 \pm 0.20$	$0.81 \pm 0.30$
$\delta_{\text{SQ}}$	$0.63 \in [0.50, 0.74]$	$0.63 \in [0.50, 0.74]$	$0.63 \in [0.50, 0.74]$
$\delta'_{\text{SQ}}$	$0.74 \in [0.43, 0.79]$	$0.74 \in [0.43, 0.79]$	$0.74 \in [0.43, 0.79]$
$\tilde{\beta} = \frac{1}{2 - \delta_{\text{SQ}}}$	$0.73 \in [0.66, 0.74]$	$0.73 \in [0.66, 0.74]$	$0.73 \in [0.66, 0.74]$
Large $T_\phi$			
$\beta$	$0.51 \pm 0.01$	$0.62 \pm 0.14$	$0.73 \pm 0.17$
$\log T_\phi >$	15	17	18
$\delta_{\text{eff}} = 2 - \frac{1}{\beta}$	$0.04 \pm 0.02$	$0.39 \pm 0.23$	$0.63 \pm 0.23$
$\alpha_{\text{eff}} = \frac{1 - \delta_{\text{eff}}}{2 - \delta_{\text{eff}}}$	$0.49 \pm 0.01$	$0.38 \pm 0.09$	$0.27 \pm 0.08$
$\alpha$	$0.52 \pm 0.02$	$0.36 \pm 0.14$	$0.44 \pm 0.13$
$\log T_\phi \in$	[16, 19]	[17, 19.8]	[15.8, 18]

to match. However, fitting the whole curve structure may be problematic in this context. Indeed, the traders with a typical small value of  $T_\phi$  see a more linear relationship in the region of small transaction value than when considering the whole curve. For instance, removing the two largest segments from the fee structure yields  $\delta'_{\text{SQ}} = 0.74 \in [0.43, 0.79]$ , which is not far from  $\delta_{\text{eff}}$ .

2. *Large values of  $T_\phi$ :* The relationships between all the exponents are verified for the three categories of clients. While not very impressive for companies and asset managers, this result is much stronger in the case of retail clients since the relative uncertainties associated with each measured exponent are small (1–2%). The value of  $\beta_{\text{retail}}$  is of particular interest as it corresponds to  $\delta_{\text{eff}} = 0$  or, equivalently, to a flat fee structure. Going back to the fee structure of Swissquote, one finds that the transition happens when the relative transaction cost falls below some threshold (we cannot give its precise value for confidentiality reasons; it is smaller than 1%). A possible explanation is that either some traders with a high enough average turnover have a flat fee agreement with Swissquote or the rest of them simply act as if they were not able to take correctly into account transaction costs. Since not all traders have a flat fee agreement, one must conclude that some traders have indeed some problems estimating small relative fees and simply disregard them. The reported value of  $\beta$  for companies and asset managers is larger than  $\beta_{\text{retail}}$ , but it is more likely that the small sample size is responsible for this discrepancy, since these two categories of clients have a greater propensity to negotiate a flat fee structure.
3. *Transition between the two regimes:* The transitions between the standard Swissquote and an effective flat fee structure occur *at the same average value* of  $T$  for the three categories of traders (idem for  $T_\phi$ ). Since there is no automatic switching between fee structures at Swissquote for any predefined value of transaction value, one is led to conclude that this transition has behavioral origins, which is also responsible for the value at which the

transition takes place, which, in passing, corresponds to the end of the plateau of the RCDF of  $P_v$  in the case of retail clients ( $e^{15} \simeq 3.27 \times 10^6$ ). As a consequence, it is likely that the traders tend to either neglect or consider as constant transaction fees smaller than some threshold when they build their portfolio.

## 6. Discussion and outlook

We have been able to determine empirically a bilinear relationship between the average log-turnover and the average log-account value, and have argued that it comes from the transaction fee structure of the broker and its perception by the agents. A theoretical derivation of optimal simple one-shot mean-variance portfolios with non-linear transaction costs predicted relationships between turnover, number of different assets in the portfolio and log-account values that could be verified empirically. This means that the populations of traders do take correctly *on average, i.e. collectively*, the transaction costs into account and act *collectively* as mean-variance equally weighted portfolio optimizers. This is not to say that each trader is a mean-variance optimizer, but that the population taken as a whole behaves as such—with differences across populations, as discussed in the previous section. This is to be related to findings of Kirman's famous work on demand and offer average curves in Marseille's fish market [39], and more generally to what has become known as the wisdom of the crowds (see [40] for an easy-to-read account).

The fact that the turnover depends in a nonlinear way on the account value implies that linking the exponents of the distributions of transaction volume, buying power of large players in financial markets and price return is more complex than was previously thought [16]. It also has implications for agent-based models, which from now on must take into account the fact that the real traders do invest in a number of assets that depend nonlinearly on their wealth.

Future research will address the relationship between account value and trading frequency, which is of utmost importance to understand if the many small trades of small investors have a comparable influence on financial market as those of institutional investors. This will give an understanding of who provides liquidity and what all the nonlinear relationships found above mean in this respect. This is also crucial in agent-based models, in which one often imposes such a relationship by hand, arbitrarily. Conversely, one will be able to validate evolutionary mechanisms of the agent-based model according to the relationship between trading frequency, turnover, number of assets and account value they achieve in their steady state.

## Acknowledgment

DMdL is grateful to Swissquote Bank for financial support.

## References

- [1] Mantegna R and Stanley H G 2000 *Introduction to Econophysics* (Cambridge: Cambridge University Press)
- [2] Bouchaud J-P and Potters M 2000 *Theory of Financial Risks* (Cambridge: Cambridge University Press)
- [3] Dacorogna M M, Gencay R, Müller U A, Olsen R B and Pictet O V 2001 *An Introduction to High-Frequency Finance* (London: Academic)
- [4] Bouchaud J-P, Farmer J D and Lillo F 2008 How markets slowly digest changes in supply and demand arXiv:0809.0822

- [5] Cont R and Bouchaud J-P 2000 Herd behaviour and aggregate fluctuation in financial markets *Macroecon. Dyn.* **4** 170
- [6] Challet D, Marsili M and Zhang Y-C 2005 *Minority Games* (Oxford: Oxford University Press)
- [7] Raffaelli G and Marsili M 2006 Dynamic instability in a phenomenological model of correlated assets *J. Stat. Mech.* **2006**(8) L08001
- [8] Brock W A and Hommes C H 1997 A rational route to randomness *Econometrica* **65** 1059–95
- [9] Challet D and Zhang Y-C 1997 Emergence of cooperation and organization in an evolutionary game *Physica A* **246** 407 (arXiv:adap-org/9708006)
- [10] Slanina F and Zhang Y-C 2001 Dynamical spin-glass-like behavior in an evolutionary game *Physica A* **289** 290–300
- [11] Jefferies P, Hart M L, Hui P M and Johnson N F 2001 From market games to real-world markets *Eur. Phys. J. B* **20** 493–502 (arXiv:cond-mat/0008387)
- [12] Challet D, Chessa A, Marsili M and Zhang Y-C 2000 From minority games to real markets *Quant. Finance* **1** 168 (arXiv:cond-mat/0011042)
- [13] Challet D and Marsili M 2003 Criticality and finite size effects in a realistic model of stock market *Phys. Rev. E* **68** 036132
- [14] Maslov S, Zhang Y-C and Marsili M 1998 Dynamical optimization theory of a diversified portfolio *Physica A* **253** 403–18
- [15] Bouchaud J-P, Giardina I and Mézard M 2001 On a universal mechanism for long ranged volatility correlations *Quant. Finance* **1** 212 (arXiv:cond-mat/0012156)
- [16] Gabaix X, Gopikrishnan P, Plerou V and Stanley H E 2003 A theory of power-law distributions in financial market fluctuations *Nature* **423** 267
- [17] Challet D and Stinchcombe R 2001 Non-constant rates and over-diffusive prices in a simple model of limit order markets *Quant. Finance* **3** 155
- [18] Farmer J D and Lillo F 2003 On the origin of power law tails in price fluctuations *Quant. Finance* **4** 7 (arXiv:cond-mat/0309416)
- [19] Lillo F 2007 Limit order placement as an utility maximization problem and the origin of power law distribution of limit order prices *Eur. Phys. J. B* **55** 453–9 (arXiv:physics/0612016)
- [20] Thurner S, Farmer J D and Geanakoplos J 2009 Leverage causes fat tails and clustered volatility (arXiv:0908.1555)
- [21] Arthur B W 1994 Inductive reasoning and bounded rationality: the El Farol problem *Am. Econ. Rev.* **84** 406–11
- [22] Lux T and Marchesi M 1999 Scaling and criticality in a stochastic multi-agent model of a financial market *Nature* **397** 498–500
- [23] Caldarelli G, Marsili M and Zhang Y-C 1997 A prototype model of stock exchange *Europhys. Lett.* **50** 479–84
- [24] Alfarano S and Lux T 2003 A minimal noise traders model with realistic time series properties *Long Memory in Economics* (Berlin: Springer)
- [25] Vaglica G, Lillo F, Moro E and Mantegna R N 2008 Scaling laws of strategic behavior and size heterogeneity in agent dynamics *Phys. Rev. E* **77** 36110
- [26] Barber B M, Lee Y-T, Liu Y-J and Odean T S 2005 Do individual day traders make money? Evidence from Taiwan SSRN eLibrary <http://ssrn.com/paper=529063>
- [27] Bauer R, Cosemans M and Eichholtz P 2007 The performance and persistence of individual investors: rational agents or tulip maniacs? [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=965810](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=965810)
- [28] Gondhi N R, Mangla V, Pochiraju B and De S 2010 Success/failure of past trades and trading behavior of investors SSRN eLibrary <http://ssrn.com/paper=1568642>
- [29] Clauset A, Shalizi C R and Newman M E J 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703
- [30] Yakovenko V M 2007 Econophysics, statistical mechanics approach to *Encyclopedia of Complexity and System Science* (Berlin: Springer) <http://refworks.springer>

- [31] Efron B and Tibshirani R J 1993 *An Introduction to the Bootstrap* (New York: Chapman and Hall)
- [32] Solomon S and Richmond P 2001 Power laws of wealth, market order volumes and market returns *Physica A* **299** 188–97
- [33] Schwarzkopf Y and Farmer J D 2010 Empirical study of the tails of mutual fund size *Phys. Rev. E* **81(6)** 066113
- [34] Schwarzkopf Y and Farmer J D 2010 What drives mutual fund asset concentration? SSRN eLibrary <http://ssrn.com/paper=1173046>
- [35] Brennan M J 1975 The optimal number of securities in a risky asset portfolio when there are fixed costs of transacting: theory and some empirical results *J. Financ. Quant. Anal.* **10** 483–96
- [36] Sharpe W F 1964 Capital asset prices: a theory of market equilibrium under conditions of risk *J. Finance* **19** 425–42
- [37] Veld C and Veld-Merkoulova Y V 2008 The risk perceptions of individual investors *J. Econ. Psychol.* **29** 226–52
- [38] Van de Venter G and Michayluk D 2009 A longitudinal study of financial risk tolerance *Technical Report* [http://www.fma.org/Reno/Papers/Longitudinal\\_Financial\\_Risk\\_Tolerance.pdf](http://www.fma.org/Reno/Papers/Longitudinal_Financial_Risk_Tolerance.pdf)
- [39] Härdle W and Kirman A 1995 Nonclassical demand: a model-free examination of price–quantity relations in the Marseille fish market *J. Econom.* **67** 227–57
- [40] Surowiecki J 2005 *The Wisdom of Crowds* (New York: Anchor Books)