Practice Makes Perfect?

The Effect of Experience on Overconfidence – Empirical Evidence from the Data Set of Budapest Marathons

By

Erika Dömötör

Submitted to

Central European University

Department of Economics

In partial fulfilment of the requirements for the degree of Master of Arts in Economics.

Supervisor: Professor Alessandro De Chiara

Budapest, Hungary

2016

CEU eTD Collection

Abstract

This thesis investigates the determinants of overconfidence, in particular, the effect of experience on the overestimation of one's actual performance. The question is relevant for many fields in economics, for example, for labor economics, contract theory and finance. The results of the Budapest Marathons from 1999 to 2014 serve as the database in which overconfident behavior is measured. In the methodology of the thesis an OLS model of overconfidence is built with the explanatory variables of gender, age, skill and experience. In line with previous works, I find that on average males are more overconfident than females, while age has a U-shaped effect with the minimum around the middle age. It is also shown that skill is negatively correlated with overconfidence, according to the Dunning-Kruger effect. The thesis contributes to the literature by enhancing the model with the effect of experience. The upcoming self-selection bias of the proxy for experience is treated with an insrumental variable method. After dealing with this endogeneity issue, the model shows no significant effect of experience on overconfidence.

CEU eTD Collection

Acknowledgements

I would like to thank first the whole Faculty and Staff of the Department of Economics for continuously supporting me in my studies and in planning my future. I am grateful for letting me doing my own way, but still keeping me on the right track.

I express a special thank to my supervisor, Alessandro De Chiara. For hours-long consultations, useful advice and great discussions, not only on the thesis, but on research and (academic) life as well. I am also grateful to Botond Kőszegi and Sergey Lychagin for their frequent feedback and hints on this project.

Finally, I would like to thank the fellow students in the MA and PhD programs for being around. They helped, motivated and impressed me in so many ways, that I cannot explain on this single page.

I dedicate all of my work to my Family, as a small compensation for the time I sacrifice to research.

CEU eTD Collection

Table of Contents

CEU eTD Collection

List of Figures

List of Tables

CEU eTD Collection

1 Introduction

A Marathon is not just two Half marathons - advises Sarolta Monspart[1], an icon of Hungarian running community. She is undoubtedly right considering that an average participant of Budapest Marathon completes the second half of the distance 26 minutes or 23 percent slower than the first half. Her advice refers to the well-known fact that for a long-distance runner steady pace implies the best possible personal result.[2] Therefore a notable slow-down in the second part of the distance might correspond to an overly ambitious choice of the starting pace. This argument motivates to employ the difference between the time results of the first and second half of the marathon race as the measure of the more general phenomenon of overconfidence.

The above cited advice and facts illustrate how smoothly a non-conventional data set can fit into a research in economics. One of the advantages of using sports data is their wide and easy availability. For this project I collected over 21,000 observations from the Budapest Marathons from 1999 to 2014. Although each of these observations contain relatively few information about individual charachteristics, there is something else unique in sports data. The results on races, themselfs, provide information about such personal attributes, which are hard to observe in other situations. For example, performance, efforts as well as motivation and personal goals are all captured, as it was exploited in Filippin and Van Ours (2012) and Allen at al. (2014). Overconfidence can be also effectively measured in sports data. Krawczyk and Wilamowki (2014) use the difference between individuals final and semi-distance results on marathon races. In this thesis I build on this measure to determine the effect of experience on overconfidence.

[1]Wikipedia: Sarolta Monspart – http://en.wikipedia.org/wiki/Sarolta_Monspart
[2]It was also justified by the physiological research of Joyner (1991).

CEU eTD Collection

From the point of this project the main contribution of Krawczyk and Wilamowski is defining widely available proxies for overconfidence. They build these measures purely on the final and semi-distance time results of marathoners. Furthermore, Krawczyk and Wilamowski (2014) have already proved the validity of these proxies by organizing a survey to measure directly the own forecast errors of marathoners' actual performance. Besides this worthy contribution Krawczyk and Wilamowski revealed several relevant factors of overconfidence. They showed empirically that male are more overconfident than female. They also concluded that age has less straightforward effect on overconfidence, but most likely its contribution is U-shaped with the minimum around middle age. Finally, they investigated the cultural aspects of overconfidence. In contrast, although Krawczyk and Wilamowski mention the role of experience and claim that the effect is expected to be negative, I am not aware of any previous work, which empirically investigates this issue. The main research question of this thesis is whether experience reduces overconfidence.

The model which is implemented in this project contains the measure of skill as a control variable. In my data set there is a large variation in performance between individuals, much larger than the improvement a runner can achieve within her career. Additionally, a systematic effect is also expected; it is assumed that lower skills lead to higher overestimation of one's actual performance, while higher skills result more accurate forecast. The phenomenon is called the Dunning-Kruger effect in psychology (Dunning, 2011). Recent empirical research on the issue is available in Feld et al. (2015), including evidence from undergraduate students' actual and forecasted grades. As main contribution they argue that their estimation is consistent, since they have found a suitable exogenous instrument for skill; the performance on a previous exam. In the current project it is feasible to show additional evidence on the Dunning-Kruger effect.

CEU eTD Collection

The analysis in this thesis extend the scope of Krawczyk and Wilamowski (2014) with the effect of experience on overconfidence. Although a negative relationship is expected, it is not proved based on my data set. In raw data an increase of overconfidence by the increase of experience can be observed. On the other hand, after controlling for other variables, such as gender, age and skill, the sign becomes ambiguous. Furthermore, in any setup I use in this project, the effect of experience on overconfidence is insignificant or at most weakly significant. Consequently, based on the data set of the Budapest Marathons, there is no reason to assume that experience can significantly decrease overconfidence.

Along with the main research question, this project gives further empirical evidence on the Dunning-Kruger effect, that is, skill and overconfidence are negatively correlated. There are two main advantages of using marathoners' data instead of the classical school performance survey by Feld at al. (2015). First, the runners on giant sport events are less likely to give attention to ranking, they are interested in their own performance instead. Therefore the setup with the marathoners' can measure individuals' overestimation of one's actual performance in absolute terms, while students and also professors might think on skills and grades on a relative scale. Second, grading systems have natural bounds. For example, in Feld et al. (2015) the maximum grade is 10. Therefore a very highly skilled student, let's say, with grade 9, cannot forecast much better result than the actual. This technical bound may explain the more accurate estimation of individuals with higher skills.

Understanding the nature of overconfidence is not just an issue for psychology, but relevant for economic research as well. For applications in finance see Bhanari and Deaves (2006) and Malmendier and Tate (2015). In contract theory, the consequences of overconfidence are described in Kőszegi (2014) and Heidhues et al. (2015). Finally, Krawczyk and Wilamowski (2014) illustrate their investigation with examples from labor economics.

CEU eTD Collection

2 Background

This thesis raises the question: how does experience affect overconfidence? The issue originally comes from psychology, but it has several applications in economics and finance as well. The aim of this project is to show empirical evidence on the factors affecting overconfidence, using a special data set from sports. In this way varios aspects are built together in this project to approach the main research question. This section presents these background pieces from the literature. First, the phenomenon of overconfidence is defined and categorized based on psychology and relevant applications are shown from economics. Second, I illustrate how previous works exploit data from sports. Finally, the papers which are supporting the methodology of the thesis are introduced.

2.1 Overconfidence in Psychology and in Economics

In their paper in the field of psychology Moore and Hailey (2008) defines and categorize overconfidence. In this project the term is used with the meaning of "overestimation of one's actual performance". Or using the more precise definition: "overconfidence is the overestimation of one's actual ability, performance, level of control, or chance of success" (Moore and Hailey, 2008).

The phenomenon of overconfidence influences individual deceisons, such as consumption or efforts, hence it is widely investigated in economics as well. Since the field of psychology and economics is more like a general concept, the issue of overconfidence is relevant for many areas. For an application in finance see, for example, Bhanari and Deaves (2006). In contract theory, the consequences of overconfident behavior are described in Kőszegi (2014).

CEU eTD Collection

Furthermore, Heidhues et al. (2015) deals with a supervisor-agent situation in a workplace environment. In their model an overconfident supervisor delegates tasks to the agents. Due to her overly optimistic beliefs about herself, external conditions, such as the ability of the agents, are not taken into account properly. The same issue may affect overconfident runners as well. It would be worth to think on how the external conditions, such as daily weather, build into their attitude. However, these considerations are out of the scope of the current project.

The application in the paper of Malmendier and Tate (2015) is the closest to the research question of this thesis. They investigate the investment decesions of overconfident CEOs. It is assumed that overconfidence is higher on average among top managers. A possible reason might be that climbing the corporate ladder is sided by more success than failures. Therefore, while an employee becomes CEO, she gets a series of positive feedback. From this insight, the expactation, that experience reduces overconfidence, can be challenged.

2.2 Using Sports Data from the Field

Data from sports are widely used in economic research, in particular in the fields where individuals' own deceisions are in the focus. This is applicable to many questions using behavioral approach, like in Filippin and Van Ours (2012), but other areas, such as labor economics, can benefit from it as well. For instance Kahn uses sports "as a Labor Market Laboratory" (Kahn, 2000). In his approach he gets data from professional sports, which have the advantage of knowing the exact match between employers and employees, as well as the compensations along the whole career of the agents. Furthermore, it is not less important that clear and direct measures of performance are available in sports, which is not the case in many other professions.

CEU eTD Collection

In sports not only the professionals', but also the amateurs' performance is widely reported on the websites of the races. It is certainly true for long-distance running, which becomes a more and more popular hobby, especially among highly skilled individuals. Notwithstanding, this way not the whole population is represented (for example, physical workers are unlikely to run for fun), the selection is in line with the applications in labor economics and in finance. While in many cases it is hard to measure certain attributes of individuals genuinely, the growing number of hobby marathoners undeliberately reveal various kinds of information about themselfs.

Allen and Dechow (2013) and Allen at al. (2014) collect data from amateur long-distance runners to provide evidence for theories in behavioral economics. They investigate prospect theory, in particular reference-dependent preferences. Finding a convincing evidence of reference-dependence is challenging when the reference point is not clear or heterogenous across individuals. Allen et al. argue that it is unlikely the case with the marathoners, since they often set their goals to round finish times. Furthermore, they find data from the runners convenient, because semi-distance results are reported at many points of the race, therefore the change in performance, and hence the adjustments in efforts can be easily tracked. The latter reasoning will show up in the methodology of the current project, as overconfidence is defined by the final and half-distance results.

CEU eTD Collection

2.3 Empirical Literature on Overconfidence

The methodology of this thesis builds on two papers. The interaction of gender and age with overconfidence is empirically analysed by Krawczyk and Wilamowski (2014), exploiting data from marathon runners. Feld et al. (2015) provide evidence on the relationship between skill and overconfidence, called the Dunning-Kruger effect. The latter shows up in psychological literature as well; see, for example, Dunning (2011). In contrast, the effect of experience on overconfidence is less investigated and this paper aims to fill this gap.

The recent working paper by Krawczyk and Wilamowski (2014) investigates the factors which determine individuals' overconfidence. They use a wide data set of marathon runners where a certain type of overconfident behavior, the overestimation of one's actual performance can be captured. The first part of their paper deals with the construction of a suitable proxy for overconfidence, in which only final and semi-distance results are used, and hence which is available for a large set of runners. In this thesis the same measure will be applied. In their paper Krawczyk and Wilamowski also raise the question on the role of experience in overconfident behavior, but do not analyze it empirically in their studies. The contribution of this paper is the enhancement of the model with the effect of experience on overconfidence.

Feld et al. investigates the so-called Dunning-Kruger effect, that is overconfidence increases with the decrease of skill. They use a classical experimental environment; they analyze university students' actual grades and their own forecasted performance. The contribution of Feld et al. (2015) is the adjustment of the bias which comes from defining both the skill and overconfidence measures by using the same actual performance. To avoid this, measurement of skill is captured from previous results. The limitations of their estimation are the following. First, the grading system has an upper limit (grade 10 in their study) which causes a technical bias. Second, grades are more or less relative in any grading system, therefore only relative overconfidence can be measured accurately. An additional contribution of this thesis is that it can avoid these concerns. First, there are no specific upper or lower bounds in the measures used in current project. Second, amateur runners'

CEU eTD Collection

results reflect individual and not relative performance, and therefore overconfidence can be defined in absolute terms.

In this paper I remake the model of Krawczyk and Wilamowski with two main demographic attributes, gender and age, and add also a skill measure following Feld et al. (2015). Furthermore, I extend the analysis with an additional factor – experience. The basic model is built by an OLS method while the endogeneity of the proxy used for experience is adjusted by an instrumental variable. The choice of the instrument, the distance the runner have to travel for participating on the sport event, is motivated by the paper of Card (2001). In his well-known contribution the proximity to schools is applied as an instrument for the deceision to take higher education.

CEU eTD Collection

3 Data and Variables

In this project I build up an own data set from recent result tables of Budapest Marathons. Similar information would be available for other running events around the world, possibly with millions of participants. On the other hand, for an adequate definition of experience it is necessary to more or less track the individuals over the years. Therefore in this analysis I stick to a single series of events, the Budapest Marathons. The final data set contains 21,839 observations from 16 years. In this section first, I describe the source and organization of the data, including special features and issues emerging from the structure. Then, I define the variables which I use during the analysis. As one of the main points of the methodology, measures for overconfidence are discussed in details. Finally, basic dependencies in the data are shown before moving on to the methodology and empirical results.

3.1 From Raw Data to Data Set

The raw data are from the webpage of Budapest Sportiroda[3], the organizer of Budapest Marathons. On the webpage detailed results are available in yearly tables, altogether for 29 events from 1984 to 2014. Time results are reported for each runner at the start line, finish line and half distance. Additionally, basic demographic information, such as gender, date of birth, nationality and city, is available. This data set has the advantage of having a huge number of observations; however, it lacks variety in terms of the variables describing individual characteristics. Furthermore, data availability changes over the time periods: it is more detailed in more recent years. Therefore restrictions are necessary before carrying out the analysis.

[3]Website: http://www.futanet.hu/cikk/10999

CEU eTD Collection

First, in the final sample only those runners who started marathon running after 1999 are considered, because half-distance results are provided only after that year. Second, only data of Hungarian participants are included in the data set. It is assumed that long-distance runners from Hungary are likely to participate frequently in Budapest Maratons since it is the only large marathon event in the country. Therefore the number of races finished previously is a promising measure of their experience. In contrast, foreign runners are more likely to participate in Budapest Marathons occasionally; hence their experience cannot be reliably measured based on the available data.

3.2 Data Set Features and Issues

On the final data set several checks for possible issues and data cleansing were performed. Runners with missing relevant data were excluded, as well as the ones where obviously wrong information, such as negative time result or age below the official limit of 18, were reported. Finally, an adequately large data set with more than 21,000 observations became available in 16 different tables from 1999 to 2014.

In the second step of building the data set yearly tables were matched. The linking of participants between the years were based on name and date of birth. This information generally well defines individuals but still gives some chance for data issues. The most obvious ones, such as clerical errors, were checked and adjusted. Moreover, it is still possible that two different participants were accidentally merged or that one participant was counted twice. It is assumed that these measurement errors occur randomly, and therefore do not affect the results.

CEU eTD Collection

The primary goal of matching yearly tables are to extract information on the runners' experience. The corresponding measure is the number of Budapest Marathons the individual participated before the current race. On the margin, the data set gets a panel form, which motivates to think on using panel data methods as well. However, due to the initial structure of the data – not everybody runs every year – the set of observations are heavily unbalanced. Because of this consideration, current project is restricted to pooled cross-sectional methods, but the structural feature of having panel form is worth to keep in mind for two reasons. First, data may contain additional information as individuals can be tracked, which could be exploited in further analysis. Second, there might be a serial correlation between the error terms. To avoid issues from the latter, clustered errors are used during the analysis. It is a conservative approach, which increases the variance of the estimates.

An additional concern arises due to the sampling method, in which only individuals who started marathon running after 1999 are considered. This way runners with less experience are more likely to get into the sample. To balance this, probability weighting is used in the regressions. It is assumed that the distribution of experience in the true population is the same as in the last available year. For example, there are only two runners in the whole sample who participated in 15 previous marathons, both observed in 2014. Therefore, there are two observations with this high experience in the last year sample of size 2,000 as well as in the full sample of size 20,000. The probability weight to adjust this sampling charachteristic is around 10.

CEU eTD Collection

3.3 Dependent Variables

In the regressions several types of dependent variables are considered. First, gender and age represents the individuals' demographic characteristics, which are naturally exogeneous. Second, the main factor of interest, experience, is measured by the number of previous Budapest Marathon participations. As more overconfident runners may participate in more events, the estimated parameter of the main variable might be biased. The distance between the home town of the particpant and the event (Budapest) performs as an instrument to avoid endogeneity. Third, some proxy of skill is also used in the regressions. In the main models the performance on the individuals' first marathon is considered, which can be interpreted as initial ability before gaining any experience.

Technically, the skill measure from the first participation is independent only for marathoners with at least one previous event experience. As skill is not the primary focus of the analysis, it does not cause major issues, but as a robustness check I run all regression on a restricted sample, where first marathoners are excluded. Unfortunately, it comes with the loss of half of the observations and furthermore, the effect of changing experience from zero to one event is missed. There are additional options to measure skill. In an alternative setup result on current event is used, possibly intrumented with the first marathon result. In further regressions the finish time is replaced by half-distance time, however these models have less explanatory power.

CEU eTD Collection

To summarize the subsection, the description of the variables is the following:

- *gender*: The gender of the participant, 1 for a male and 0 for a female.
- *age*: The age of the participant in the year of the marathon.
- *count*: Number of Budapest Marathons previously finished.

- *distance*: The distance of the participants' home town from the event (in kilometers).[4]

- *finish*: Net absolute time result (in seconds).[5]

- *finishfirst*: Net absolute time result (in seconds) from the earliest reported race of the participant.

- *half*: Net time result (in seconds) at half distance.

- *halffirst*: Net time result (in seconds) at half distance from the earliest reported race of the participant.

The listed variables are all available or can be extracted from the raw data. Additionally, a suitable proxy for overconfidence have to be found. The next subsection shows how it is constructed from the above information, and also verifies its validity.

3.4 The Measures of Overconfidence

Krawczyk and Wlilamowski (2014) initially defined overconfidence with the Absolute Forecast Error (AFE) or the Relative Forecast Error (RFE). AFE and RFE are both direct measures of overconfidence; they are calculated from the actual result and the pre-race forecasted result by the runner. Therefore, AFE and RFE directly indicate the overestimation of the runners' actual performance. Unfortunately, this information can be collected only by surveying indviduals about their estimated result, which is not feasible in this project. On the top of that, to get reliable data, self-reporting of the own forecasts should be incentivized.

[4]This variable was defined based on information collected from Google Maps. The length of the shortest route by car was used.

[5]In the raw data the 'clock' time at finish and start line is available. The actual result is the difference of these two. Using the term 'net' reflects to this correction.

CEU eTD Collection

In their study Krawczyk and Wilamowski organized a survey to collect the runners' own forecasts before the race. Reporting true valuations was motivated with a forecasting game, where runners could get a gift for tipping their final time result accurately. To avoid 'cheating' by reporting a lower performance and decrease efforts on the race, the prize was smartly designed in such way that the runners' own forecasted time was indicated; it was printed on a T-shirt. Winning the T-shirt is clearly a gain, but showing on it a worse result than the expected performance is a loss for the runner. The size of the collected data set was moderate compared to the number of marathon participants but it was adequate to find and validate two widely available proxies.

The measure which can be used for a larger set of available data is based on the well-known fact that to achieve the ideal result, it is the best strategy to keep a steady pace during the race. It is proved in sports literature as well; such as in Joyner (1991). Moreover, online sources for marathoners send the same message:

"...start even a few seconds per mile too fast, and misery awaits: excess fatigue, loss of motivation, or even injury... At a race, you'll get the best results if you try to maintain a steady pace from start to finish." [6]

In their survey Krawczyk and Wilamowski showed empirical evidence that the failure in keeping the steady pace is closely correlated with the overestimation of one's actual performance. In practice, a steady pace means that the time result on the second half of the distance is close to the time result on the first half. The larger is the gap, the more overconfident the runner was, when chose the pace at the beginning of the race.

Based on the above considerations, the new measures of overconfidence become the Absolute Pace Change (APC) and the Relative Pace Change (RPC), defined in the following

[6] Runner's World: Your Perfect Pace – http://www.runnersworld.com/running-tips/your-perfect-running-pace

CEU eTD Collection

way:

$$APC = finish - 2 \cdot half$$

$$RPC = \frac{finish - half}{half} * 100$$

The first measure, APC is verbally described as *how much the runner was slower in the second half compared to the first half (in seconds).* While $RPC - 100$ shows *how much percent the second half takes longer than the first half.* In both cases the higher variable reflects the higher overconfidence. Using APC has the advantage that it is easier to interpret, as it directly shows increments in time in seconds. On the other hand, the relative definition of RPC provides more general results, less connected to actual performance. In this thesis all regressions are run for both APC and RPC, but the results are usually pairwise similar owing to the fact that the two proxies measure basically the same object.

3.5 Basic Dependencies in the Data

The summary of the variables used in this project can be found in Table 1.

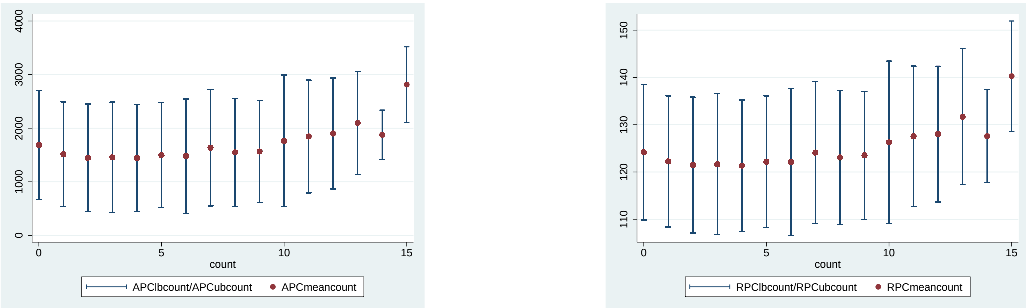
Variable	Mean	Std. Dev.	Min	Max
$gender$	0.80	0.40	0	1
age	36.10	9.08	18	79
$count$	1.33	2.00	0	15
$finish$	15178.29	2169.90	8377	21730
$finishfirst$	15333.99	2043.11	8377	21730
$half$	6794.90	808.95	4064	9970
$halffirst$	6877.85	787.24	4064	9970
APC	1588.49	1013.31	-4270	6633
RPC	123.10	14.34	55.33	220.88

Table 1: Summary of the Variables

The majority, 80% of the runners are male and the average age is 36.10. The average finish result is 15,178 seconds, which is around 4 hours and 13 minutes. On the other hand, the

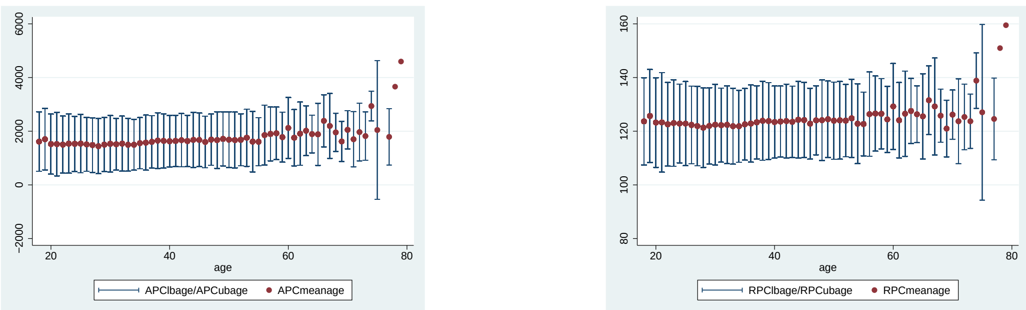
CEU eTD Collection

average half-distance result is 6,795 seconds, which is about 113 minutes, a little bit less than two hours. So far, based on APC and RPC measures, it seems that the population is generally overconfident. An average runner is more than 26 minutes (1,588 seconds) or 23 percent slower in the second half of the distance. Less experienced runners are in large proportion in the true population, and on the top of that, they are overrepresented due to the sampling method. However, there are also a few very experienced runners who participated in almost all Budapest Marathons during the 16 years. Finally, the mean of the experience measure, *count* is low (individuals participated only on 1.33 previous events on average), while the standard deviation compared to this is high (2.00).



(a) APC (b) RPC

Figure 1: Relationship between Experience and Overconfidence

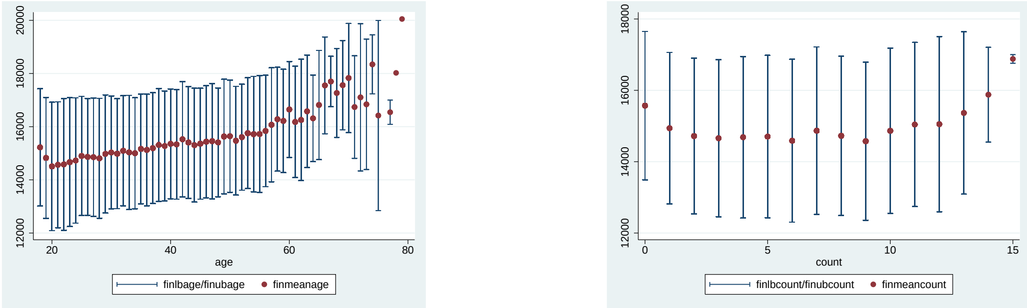


(a) APC (b) RPC

Figure 2: Relationship between Age and Overconfidence

Basic dependencies in the data are shown on graphs. First, Figure 1. illustrates the relationship between experience (*count*) and measures of overconfidence (APC or RPC). A

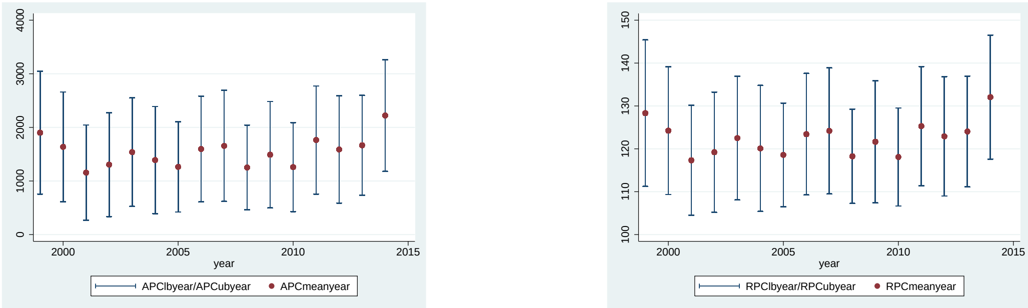
CEU eTD Collection



(a) Age - Result

(b) Experience - Result

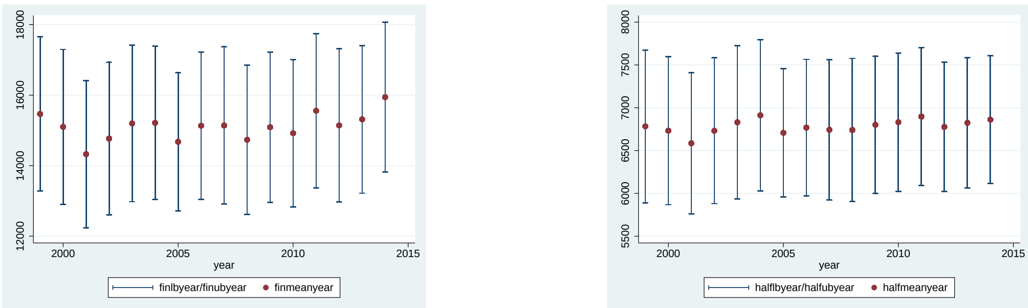
Figure 3: Relationship between Age, Axperience and Result



(a) APC

(b) RPC

Figure 4: The Yearly Variation in Overconfidence



(a) finish

(b) half

Figure 5: The Yearly Variation in Performance

slight increasing trend can be observed. However, looking also for non-linear forms, one can argue that the first few years of experience decrease, while additionals are increase overconfidence. Beyond this basic relationship, there are many other factors affecting the measure of overconfidence, which should also be taken into account.

First, *age* has a very similar effect on overconfidence as *count*, as Figure 2 shows. It is also true that *age* and *count* are strongly correlated, owing to the simple fact that gaining

CEU eTD Collection

experience takes time, as Budapest Marathon is organized only once a year. Based on the strong correlation it makes sense to run models without the variable *age*.

Second, both age and experience affects the final result as well. Figure 3. illustrates these relationships. Finish time increases with age, while the effect of experience shows similarities to the effect observed regarding overconfidence.

Finally, there are notable differences in both average finish times and average overconfidence measures between the years. For example, on the race of 2014 results were worse than on other events. A possible explanation is the change in external conditions; in 2014 the weather was much warmer than usually, which strongly affects the performance in running. To control for these devitions yearly dummies are also included into the models. The yearly variation in average overconfidence is illustrated on Figure 4., while Figure 5. shows the variations in the average final and semi-distance results. On the latter graphic note that the half-distance performance changes much less across the years. The differences in the final results, which are mightly due to the external conditions, almost entirely build into the overconfidence measure.

CEU eTD Collection

4 Methodology

The methodology consists of two parts. First, the effect of experience on overconfidence is investigated by an OLS model with variables of basic characteristics, *gender*, *age*, a proxy for skill, yearly dummies and the measure of experience. Second, the endogeneity of the variable of main interest, experience, is treated with an instrument. The results are reported both for *APC* and *RPC*, and as a robustness check, for different model setups.

4.1 Main Model of the Effect of Experience on Overconfidence

In the first part of the project a basic OLS model is built to explain *APC* with the demographic factors *gender* and *age* as well as with some measures for skill and experience. Additionally, yearly dummies ware added to control for external conditions in each year. Gender is a natural explanatory variable; it is stated and proved many times (for example in Krawczyk and Wlilamowski, 2014 that males are more overconfident than females. For *age* a U-shaped effect is expected with the minimum around middle age, but this effect is not be clearly seen in the data. On the other hand, adding the square of age to the regressions usually does not change the effect of the main variable, and variables *age* and *agesq* are jointly significant in most cases.

CEU eTD Collection

A less obvious point of the model is how to add skill measure to the model. There is a natural large variation between individuals in terms of marathon performance; therefore using related variables increase the explanatory power of the model. Furthermore, it is a fact known from psychology that skill decreases overconfidence, called the Dunning-Krueger effect (see for example Dunning (2011) or Feld et al. (2015)). In the main models the time result on first marathon ($finishfirst$) is used, which can be interpreted as general ability in running. An alternative is to measure the current skill with actual result

($finish$) and use $finishfirst$ as an instrument to avoid endogeneity issues. There are several other setups, for example, results at half distance also reflects ability. Therefore, analogues of $finishfirst$ and $finish$, $halffirst$ and $half$ might be also feasible measures of skill.

The most suitable model is the following in general form:

$$Overconfidence = \alpha_0+\alpha_1 gender+\alpha_2 age+\alpha_3 age^2+\beta_1 Skill+\beta_2 Experience+\sum \gamma_t year_t+u$$

The proxy for $Overconfidence$ is APC (or RPC). The measure of $Skill$ is $finishfirst$ in the main model, while $Experience$ is defined by $count$. Adding these variables to the general equation, the model becomes:

$$APC/RPC = \alpha_0+\alpha_1 gender+\alpha_2 age+\alpha_3 age^2+\beta_1 finishfirst+\beta_2 count+\sum \gamma_t year_t+u$$

Based on the previos literature, the following results are expected:

- $gender$ – The effect on $Overconfidence$ is positive and significant (men are more overconfident than women).
- age – The function of the effect is U-shaped with the minimum at middle age (35-45 years).
- $Skill$ – Negatively correlated with $Overconfidence$. In this setup, the lower performance indicator (time result) corresponds to higher skill, therefore the estimated parameter in the model should be positive.
- $Experience$ – Intuitively, a negative effect on $Overconfidence$ is expected.

The non-negative estimate of the effect of experience from overconfidence might coming

CEU eTD Collection

from an endogeneity issue. That is, initially more overconfident runners are tend to participate on more events and hence gain more experience. The systematic error from omitting the true measure of initial $Overconfidence$ causes an upward bias, which is handled with an exogenous instrument.

4.2 Endogeneity in the Effect of Experience

General experience in running is hard to measure since results are available only for a few larger races, but not for small events or trainings. In the current approach *count* is used as a proxy for experience, however an issue with this is the possibility of self selection bias. It was intended to show that *count* has an effect on APC but one can imagine that it is the other way around. Specifically, more overconfident runners may enter more races, therefore they gain more experience. Consequently, the omitted variable, $Overconfidence$ in a general sense has positive correlation with *count*. As the proxy APC is naturally positively correlated with $Overconfidence$ as well, the bias of *count* is anticipated to be positive, assuming that there is no additional correlation between the explanatory variables.

To avoid the endogeneity issue of *count*, an instrumental variable method is introduced. The instrument used for *count* is *distance*. There are no solid reasons to assume that the distance of the participant's home town from the event has a direct effect on APC, therefore the instrument is considered to be exogeneous. At the same time it is correlated with *count* since travelling too long may prevent runners from participating in the Budapest Marathons. Consequently the instrument is assumed to be relevant. Although the instrument seems weak based on simple correlation, its statistical significance is proved by the first-stage estimation. Additionally, the weakness is partly compensated by the large sample size.

CEU eTD Collection

4.3 Robustness Check and Alternative Setups

A robustness check deals with the concern that the measure for skill is not fully independent. For those who participate on their first marathon the final result is used for both measuring initial running ability and overconfidence. Excluding them from the sample leaves a still adequately large data set with 11,115 observations. Based on the assumption that experience is not correlated with skills, results are expected to be qualitively similar to the main model. However, if the effect of experience on overconfidence is not linear, then loosing information on the difference between first and second marathoners would end up in significant shortness of the model.

A first set of alternative models varies in the proxy for skill. Instead of the first marathon result ($finishfirst$) the current result is used ($finish$). To avoid endogeneity, $finishfirst$ can be applied as an instrument. A more different approach is to replace the final results with half-distance results. In additional setups $halffirst$ acts as skill measure. Moreover, the variables of age, which have less explanatory power in the model may be dropped.

Finally, non-linear models are also tested to challange that experience has constant effect over time. It is more reasonable to think that the first few participations can reduce more overconfidence than the laters. In an alternative model the square of count is added to the regressions. The caveat of using the more sophisticated non-linear models is that combining them with the weak instrument, *distance*, they usually become insignificant.

CEU eTD Collection

5 Empirical Results

In this section discussion of all described setups is presented. The results of the main model are shown in Table 2. While tables for the robustness check and alternative setups can be found in the Appendix.

5.1 Main Models

In the models the proxies of overconfidence, APC or RPC, are run on $gender$, age, $agesq$, $count$, $finishfirst$, yearly dummies and $count$, as the variable for experience, which is instrumented with $distance$. Results are shown in Table 2. The effects of yearly dummies are hided due to the lack of space, but they are jointly significant. The order of the models in columns are the following:

(1) Regression of APC without the instrument for $count$.

(2) Regression of APC with the instrument for $count$.

(3) Regression of RPC without the instrument for $count$.

(4) Regression of RPC with the instrument for $count$.

The effect of gender is highly signifcant in all cases, that is, men are on average more overconfident than women. In magnitude it means around 390 to 420 seconds (6.5-7 minutes) or 6.6 to 7.2 percent points of the first half of the distance. Age variables, age and $agesq$ are jointly significant in all models. They show the expected U-shape effect with minimums between 29 and 42 years. Skill measures are also in line with the preliminary expectations, they have significant positive effect. The magnitude in case of APC is between 0.25 and 0.27, which is high in practical terms as well. It means that at least one

CEU eTD Collection

quarter of the final time result builds into the overconfidence measure. The main explanatory variable, *count* has insignificant effect. Consequently, there is no clear evidence in the data that experience can significantly decrease overconfidence.

In Table 3. in the Appendix all setups of the main model are re-run for the restricted sample, where the first marathoners are excluded. In this robustness check the previously reported findings qualitatively does not change. Since this latter setup provides independent skill measures for all observations, it is worth to mention that, the effect of skill remains positive and significant. Therefore, the evidence on correlation between skill and overconfidence holds in this case as well.

	(1) APC	(2) APC	(3) RPC	(4) RPC
gender	392.3***	422.6***	6.60***	7.22***
	(16.70)	(32.21)	(0.24)	(0.55)
age	-31.57***	-34.68***	-0.43***	-0.49***
	(6.01)	(7.70)	(0.089)	(0.13)
agesq	0.40***	0.56***	0.0050***	0.0084***
	(0.083)	(0.16)	(0.0012)	(0.0027)
count	0.47	-115.5	0.091	-2.28
	(6.20)	(89.69)	(0.086)	(1.51)
finishfirst	0.27***	0.25***	0.0031***	0.0027***
	(0.0043)	(0.016)	(0.000061)	(0.00027)
Constant	-1,724***	-1,346***	87.53***	95.27***
	(134.5)	(337.2)	(1.99)	(5.71)
Obs.	21,839	21,839	21,839	21,839
R-squared	0.37	0.31	0.28	0.14

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

Table 2: Summary of Full-sample Models

CEU eTD Collection

5.2 Alternative Setups

Finally, several alternative setups are performed to strengthen the results of the main models. First, the measure of skill is varied. Instead of the first marathon result either the current result or the half-distance result is used. Second, age variables are excluded from the model, since they have only weak explanatory power. Finally, a non-linear model with adding the square of experience is tested.

In Table 4. and 5. $finish$ was used as the proxy for skill. To avoid endogeneity a variations with $finishfirst$ as an instrument are also run. The order of the models in columns are the following:

(1) Variation of the main model using $finish$ as the proxy for skill.

(2) Variation of the main model using $finish$ instrumented with $finishfirst$.

(3) The same as setup (1) with *count* instrumented by *distance*.

(4) The same as setup (2) with *count* instrumented by *distance*.

In this version of the model the effect of experience on overconfidence become significant. Without the instrument for *count* it is postive, however practically small. Additional one year of experience increase the measure of overconfidence by around 20 second or 0.3 percent points on average. With the instrument *distance* for *count* the effect is negative, but only weakly significant. On the other hand, the decrease in overconfidence is practically high in this case. It is over 220 second (more than 3.5 minutes) or 3.4 percent points. Despite the large effects, since the standard errors are much higher, these setups are still unable to prove convincingly the decrease in overconfidence as due to gaining experience.

CEU eTD Collection

Models in Table 6. use $halffirst$ as the proxy for skill, otherwise the setups are analogous to the main model in Table 2. The explanatory power of these models are less, the R-suared values are significantly dropped. The results in these setups are less stable to other changes, such as adding the instrument for *distance*. In terms of the main variable of interest, *count*, there is no qualitative change in the parameters; it generally has an insignificant effect. In Table 7. results from the models without age variables are tested, but again, it does not show any notable difference compared to the main model.

The result of non-linear models in Table 8. are in line with the expectation that overconfidence decreases with gaining experience in the first few years, but decreases later. The sign of the effect changes around 7-8 previous participations. The variables *count* and *countsq* in this model are jointly significant. The shortness of these setups is the lack of the instrument *distance*. Using instrumental variable method in this case resulted insignificant parameters.

CEU eTD Collection

6 Conclusions

In this thesis the effect on experience of overconfidence was investigated. Data set for the empirical analysis was captured from the results of Budapest Marathons in the years between 1999 and 2014. Proxies for overconfidence were provided by the gap between the time result on the first and second half of the distance of each runner. The experience of the participants were captured with the number of their previous Budapest Maratons. The regressions used additional explanatory variables, such as the demographic characteristics, gender and age, as well as a measure for skill, the result on the first marathon of the participant.

The results regarding personal characteristics, gender and age, are in line with previous literature. It was shown that males are more overconfident than females on average. Age has an U-shaped effect, with the minimum at middle age, in the main model of this project, between 29 and 42 years. The results proves also the Dunning-Kruger effect, that is, lower skill relates to higher overconfidence.

In terms of the main variable of interest, experience, none of the models showed convincingly significant effect on overconfidence. Therefore, it was not proved that experience can reduce overconfidence. This finding may reflect limitations of the data set or the methodology. First, the data are not completely homogeneous, due to the sampling method and structure. Second, although the observations are matched in a panel form, this feature was not exploited.

Besides the shortness in data and methodology, the result inspires further thinking on the reasons, why experience cannot decrease overconfidence. A first explanation is that

CEU eTD Collection

individuals do not learn from their mistakes. On the other hand, it is also possible that they do take into account the lessons from their own failures, but not aware of other external conditions. For example, while gaining experience individuals become older, and hence, on average, slower runners. If they do not adjust their expectations on performance accordingly, they tend to show increasing overconfidence. Another aspect was mentioned, but not investigated in this project. It was argued that less ideal environment, such as warmer weather, can also significantly raise the level of average overconfidence. One can think on this as an additional example, where individuals do not take into account the external conditions. These alternatives would be worth to challange empirically in the future.

CEU eTD Collection

Appendix

A Robustness Check and Alternative Setups

	(1)	(2)	(3)	(4)
	APC	APC	RPC	RPC
gender	256.5***	237.8***	4.80***	4.82***
	(28.18)	(45.58)	(0.38)	(0.59)
age	-33.33***	-33.49***	-0.41***	-0.41***
	(9.11)	(9.37)	(0.14)	(0.14)
agesq	0.48***	0.42***	0.0056***	0.0057***
	(0.12)	(0.16)	(0.0017)	(0.0022)
count	3.31	65.06	0.11	0.037
	(7.57)	(113.2)	(0.11)	(1.52)
finishfirst	0.20***	0.21***	0.0022***	0.0022***
	(0.0068)	(0.015)	(0.000096)	(0.0020)
Constant	-588.1***	-809.4*	100.4***	100.7***
	(214.4)	(450.5)	(3.18)	(6.04)
Observations	11,115	11,115	11,115	11,115
R-squared	0.24	0.22	0.19	0.19

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

Table 3: Summary of Restricted-sample Models

CEU eTD Collection

VARIABLES	(1) APC	(2) APC	(3) APC	(4) APC
gender	477.0***	451.9***	543.1***	505.3***
	(13.52)	(13.39)	(48.31)	(40.85)
age	-13.58***	-14.11***	-24.90**	-22.97**
	(4.714)	(4.671)	(11.43)	(9.509)
agesq	0.0839	0.107*	0.572**	0.489**
	(0.0632)	(0.0629)	(0.262)	(0.215)
finish	0.342***	0.315***	0.283***	0.270***
	(0.00323)	(0.00381)	(0.0288)	(0.0240)
count	23.52***	19.02***	-282.0**	-220.7**
	(4.031)	(3.978)	(135.7)	(112.3)
Constant	-3,244***	-2,804***	-2,030***	-1,877***
	(103.7)	(110.6)	(605.8)	(504.5)
Observations	21,839	21,839	21,839	21,839
R-squared	0.574	0.571	0.134	0.290

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

Table 4: Additional Full-sample Models – APC

VARIABLES	(1) RPC	(2) RPC	(3) RPC	(4) RPC
gender	7.700***	7.278***	8.740***	8.103***
	(0.211)	(0.208)	(0.764)	(0.636)
age	-0.219***	-0.228***	-0.397**	-0.365**
	(0.0776)	(0.0769)	(0.182)	(0.150)
agesq	0.00133	0.00172*	0.00902**	0.00762**
	(0.00104)	(0.00103)	(0.00414)	(0.00336)
finish	0.00402***	0.00357***	0.00308***	0.00287***
	(5.04e-05)	(6.06e-05)	(0.000453)	(0.000374)
count	0.376***	0.300***	-4.435**	-3.401*
	(0.0673)	(0.0665)	(2.147)	(1.755)
Constant	67.92***	75.32***	87.05***	89.63***
	(1.690)	(1.816)	(9.556)	(7.877)
Observations	21,839	21,839	21,839	21,839
R-squared	0.429	0.425		0.091

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

Table 5: Additional Full-sample Models – RPC

CEU eTD Collection

	(1) APC	(2) APC	(3) RPC	(4) RPC
gender	364.6***	321.6***	5.57***	5.03***
	(20.12)	(41.74)	(0.27)	(0.55)
age	-26.69***	-20.28**	-0.35***	-0.27**
	(6.26)	(9.18)	(0.090)	(0.13)
agesq	0.38***	0.034	0.0048***	0.00040
	(0.084)	(0.23)	(0.0012)	(0.0031)
count	-17.11***	221.3	-0.17*	2.87
	(6.63)	(139.3)	(0.092)	(1.88)
halffirst	0.38***	0.47***	0.0029***	0.0042***
	(0.012)	(0.061)	(0.00017)	(0.00082)
Constant	-178.3	-1,023*	113.8***	103.1***
	(145.9)	(536.5)	(2.15)	(7.26)
Obs.	21,839	21,839	21,839	21,839
R-squared	0.17	-	0.13	-

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

Table 6: Models with Half-distance Skill Measure

	(1) APC	(2) APC	(3) RPC	(4) RPC
gender	391.2***	436.6***	6.57***	7.48***
	(16.97)	(43.25)	(0.24)	(0.74)
count	0.86	-121.1	0.062	-2.38
	(6.16)	(100.6)	(0.085)	(1.71)
finishfirst	0.27***	0.25***	0.0030***	0.0027***
	(0.0041)	(0.013)	(0.000059)	(0.00029)
Constant	-2,309***	-1,855***	79.47***	88.57***
	(73.74)	(388.9)	(1.05)	(6.68)
Obs.	21,839	21,839	21,839	21,839
R-squared	0.37	0.29		

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

Table 7: Models without Age Variables

CEU eTD Collection

VARIABLES	(1) APC	(2) RPC
gender	381.7***	6.458***
	(17.07)	(0.242)
age	-21.12***	-0.291***
	(6.094)	(0.0906)
agesq	0.284***	0.00356***
	(0.0841)	(0.00124)
count	-54.88***	-0.682***
	(8.754)	(0.125)
countsq	7.269***	0.101***
	(1.219)	(0.017)
finishfirst	0.279***	0.00319***
	(0.00446)	(0.000063)
Constant	-2,589***	74.95***
	(131.1)	(1.937)
Observations	21,839	21,839
R-squared	0.316	0.215

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

Table 8: Summary of Non-linear Models

CEU eTD Collection

References

ALLEN, E.J. and DECHOW, P.M., 2013. The 'Rationality' of the Long Distance Runner: Prospect Theory and the Marathon. *Available at SSRN 2342396.*

ALLEN, E.J., DECHOW, P.M., POPE, D.G. and WU, G., 2014. Reference-Dependent Preferences: Evidence from Marathon Runners. *Working paper.*

BHANDARI, G. and DEAVES, R., 2006. The demographics of overconfidence. *The Journal of Behavioral Finance*, 7(1), pp. 5-11.

CARD, D., Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5), 1127-1160.

DUNNING, D., 2011. The Dunning-Kruger Effect: On Being Ignorant of Of One's Own Ignorance. *Advances in Experimental Social Psychology*, 44.

FELD, J., SAUERMANN, J. and DE GRIP, A., 2015. Estimating the Relationship between Skill and Overconfidence. *Working Paper.*

FILIPPIN, A. and VAN OURS, J. C.,2012. Run for fun: Intrinsic motivation and physical performance. *Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit*, No. 6387

HEIDHUES, P., KŐSZEGI B. and STRACK P., 2015. Unrealistic Expectations and Misguided Learning. *Working Paper.*

JOYNER, M.J., 1991. Modeling: optimal marathon performance on the basis of physiological factors. *J Appl Physiol*, 70(2), pp. 683-687.

KAHN, L.M., 2000. The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14, 75–94.

KŐSZEGI, B., 2014. Behavioral Contract Theory. *Journal of Economic Literature*, 52(4), 1075-1118.

KRAWCZYK, M., WILAMOWSKI M., 2014. Are we all overconfident in the long run? Evidence from one million marathon participants. *Working paper.*

MALMENDIER, U. and TATE, G., 2015. Behavioral CEOs: The Role of Managerial Overconfidence. *Journal of Economic Perspectives*, 29(4), 37-60.

MOORE, D. A., HEALY P. J., 2008. The Trouble with Overconfidence. *Psychological Review*, 115(2), 502-517.

CEU eTD Collection