

The Survival of Mediocre Superstars in the Labor Market

Replication archive

Thomas Peeters, Stefan Szymanski, and Marko Terviö

Journal of Law, Economics and Organization

Introduction

This file contains the instructions for replicating the analysis in the paper “The Survival of Mediocre Superstars in the Labor Market” by Thomas Peeters, Stefan Szymanski, and Marko Terviö. The replication archive is permanently stored under the following doi: 10.25397/eur.16692070. We first detail the source datafiles, the variables they contain, and their sources. These files are stored as Stata dta-files. Then we explain the separate steps in the analysis with their accompanying code files. The code is written in Stata 17 do-files. Along with the code we discuss the main output tables and graphs and explain where they appear in the manuscript. Please note that replicating this analysis, in particular step 1, may take a large amount of CPU time. We therefore provide the intermediate files resulting from this step. The code also allows to split the analysis in several parts which can be run in parallel on several processors.

1. Source data files

The main source data for the analysis is the file `gameleveldata.dta`. This file is at the game-club level, implying that each game leads to 2 observations, one from the perspective of each club. The file contains all games played in the top 4 English football divisions in the seasons 1973-74 through 2010-11. This is a total of 154,434 observations for 77,217 independent games. The variables refer to a game g , played in season t , from the perspective of club c , employing manager m . The file contains the variables detailed below, where the prefix `op_` is added to indicate the variable pertains to the opposing club. A star indicates that the variable is not available with the `op_` prefix.

1.1. Club-season and financial variables

VARIABLE	DEFINITION	OBS.	TYPE
Team	Name of club, anonymized to guarantee manager anonymity	154,434	String
Id	Number identifier for club	154,434	Count
Numberman	Cumulative number of managers at club c until time t	154,434	Count
Clycumman	Cumulative number of managers at club c in season t	154,434	Count
Division	Division club c plays in season t	154,434	Count

Rpi*	UK retail price index at June 30 th of year t	154,388	Continuous
Rpimulti*	Inverted RPI for multiplication for year t	154,388	Continuous
Lagrpimulti*	Inverted RPI for multiplication for year t-1	154,388	Continuous
Club	Club c name in financial dataset anonymized to guarantee manager anonymity	154,388	String
Notseasacyear	Club c accounting year ends at end of season = 0, 1 otherwise	154,388	Indicator
Pos	Club c's ranking position in division at end of season t	154,388	Count
Played	Number of games club c played in season t	154,388	Count
Seasonpoints	Number of points club c obtained in season t	154,388	Count
Prom	Club c promoted going from t-1 to t = 1, 0 otherwise	154,388	Indicator
Lagprom	Club c promoted going from t-2 to t-1 = 1, 0 otherwise	154,388	Indicator
Releg	Club c relegated going from t-1 to t = 1, 0 otherwise	154,388	Indicator
Lagrel	Club c relegated going from t-2 to t-1 = 1, 0 otherwise	154,388	Indicator
Rank	Club c rank position over all division at end of season t	154,388	Count
Rev	Club c revenues in current year GBP over season t	133,478	Continuous
Emp	Club c total number of employees in season t	94,120	Continuous
Wage	Club c total labor cost incl. tax and social security in current year GBP over season t	130,116	Continuous
Lagwage	Club c total labor cost incl. tax and social security in current year GBP over season t-1	128,180	Continuous
Wagerev	Club c wage to revenue ratio (wage/rev) in year t	128,180	Continuous
Profbt	Club c profit before taxes in current year GBP over season t	135,834	Continuous
Fixass	Club c fixed assets in current year GBP over season t	147,472	Continuous
Totass	Club c total assets in current year GBP over season t	147,426	Continuous
Netass	Club c net assets in current year GBP over season t	147,334	Continuous
Insoldum	Club c suffers insolvency in year t = 1, 0 otherwise	154,388	Indicator

1.2. Game variables

VARIABLE	DEFINITION	OBS.	TYPE
Year*	Year season t ends	154,434	Count
Yearstart*	Year season t starts	154,434	Count
Match_id*	Number indicator for each game	154,434	Count
Home*	1 = club c plays at home, 0 otherwise	154,434	Indicator
Sc	Goals scored by club c in game g	154,434	Count
Goals*	Total goals scored by both clubs in gam g	154,434	Count
Letres*	H = home win, A = away win, D = draw	154,434	String
Res*	Result game g: 2 = win club c, 1 = draw, 0 = loss club c	154,434	Count
Win	Result game g: 1 = win club c, 0 otherwise	154,434	Indicator
Draw	Result game g: 1 = draw, 0 otherwise	154,434	Indicator
Loss	Result game g: 1 = loss club c, 0 otherwise	154,434	Indicator
Points	Points obtained in game g by club c, win = 3 points	154,434	Count
Elostart	Elo value club c before start game g	154,434	Continuous

Elodif	Difference in elo value club c after vs. before game g	154,434	Continuous
Elowinperc	Implied win probability based on elo ratings before game g	154,434	Continuous
Date*	Date of game in stata format	154,434	Continuous
Dayofweek*	Day of week indicator for game date in stata format	154,434	Count
Enddate*	Last date in sample	154,434	Constant
Startdate*	Earliest date in sample	154,434	Constant
Month*	Calendar month in which game g takes place	154,434	Count
Yearreal	Calendar year in which game g takes place	154,434	Count
Period*	Monthly period in sample in which game g takes place. August 1973 = 1, June 2011 = 380.	154,434	Count
Periodstartdate*	Date on which period starts in Stata format	154,434	Continuous
Periodenddate*	Date on which period ends in Stata format	154,434	Continuous

1.3. Manager variables

VARIABLE	DEFINITION	OBS.	TYPE
Manager	Name of manager, anonymized	154,434	String
Manidfe	Unique indicator for manager m across spells	154,434	Count
Manestid	Unique indicator for manager m used in estimation as it separates among caretaker managers	154,434	Count
Mannationality	Manager m nationality, UK split in sub-countries	152,784	String
Manexpbefore	Manager m working experience in seasons as manager before entry in sample	152,784	Count
Manuk	Manager m nationality UK = 1, 0 otherwise	152,784	Indicator
Manukeire	Manager m nationality UK or Ireland = 1, 0 otherwise	152,784	Indicator
Manforeign	Manager m nationality not UK or Ireland = 1, 0 otherwise	152,784	Indicator
Manexpabroad	Manager m worked outside England = 1, 0 otherwise	152,784	Indicator
Manseasabroad	Number seasons manager m worked outside England	152,784	Count
Manexpnotukeire	Manager m worked outside UK, Ireland = 1, 0 otherwise	152,784	Indicator
Manseasnotukeire	Number seasons manager m worked outside UK, Ireland	152,784	Count
Manposition	Manager m playing position in playing career	152,784	String
Mangoalkeeper	Manager m was goalkeeper = 1, 0 otherwise.	152,784	Indicator
Mandefender	Manager m was defender = 1, 0 otherwise.	152,784	Indicator
Manmidfield	Manager m was midfield = 1, 0 otherwise.	152,784	Indicator
Manforward	Manager m was forward = 1, 0 otherwise.	152,784	Indicator
Manclubapp	Games played as player by manager m	151,281	Continuous
Manplayedprof	Manager m played professional = 1, 0 otherwise	152,783	Indicator
Mannumberclubs	Number clubs manager m played for	152,783	Count
Manexpint	Manager m played for national team = 1, 0 otherwise	152,783	Indicator

Manintapp	Number games manager m played for national teams	152,783	Count
Manseascareer	Number seasons manager m has played	152,753	Count
Manplayabroad	Manager m played abroad (outside own country) = 1, 0 otherwise	152,783	Indicator
Manplaybigfour	Manager m played in big 4 (tier 1 in ENG, SPA, ITA, GER) = 1, 0 otherwise	152,783	Indicator
Mannumbseasbigfour	Number seasons manager m played in big four leagues	152,783	Count
Manpresentfinal	Manager m present in estimation sample = 1, 0 otherwise	154,434	Indicator
Manretired	Manager m retired after last observation in sample = 1, 0 otherwise	152,784	Indicator
Mancontdiv	Division manager continued in after last observation in sample. 1 = Eng tier 1 + big4, 2 = Eng tier 2 or other abroad, 3 = Eng tier 3, 4 = Eng tier 4, 5 = Eng lower.	152,784	Count
Mancontbig5	Manager m continued in big 4 or Eng tier 1 after last observation in sample = 1, 0 otherwise	152,784	Indicator
Mancontabroad	Manager m continued abroad after last observation in sample = 1, 0 otherwise	152,784	Indicator
Manothermanexp	Manager has other football management experience (scout, assistant etc) prior to entry = 1, 0 otherwise	152,700	Indicator
Manentryothermanexp	Manager has other football management experience prior to entry + spell is entry spell = 1, 0 otherwise	154,434	Indicator
Maninternhire	Manager hired internally in 1 st job = 1, 0 otherwise	152,701	Indicator
Manentryinternhire	Manager hired internally in 1 st job + spell is entry spell = 1, 0 otherwise	154,434	Indicator
Manplayermanager	Manager hired as player-manager in 1 st job = 1, 0 otherwise	152,701	Indicator
Manexplayer*	Manager m is former player of club c	154,434	Indicator
Manplayengland	Manager m played in England = 1, 0 otherwise	152,750	Indicator
Manplaycarendyear	Year of final game manager m	143,064	Count
Manengteam1-10*	Name of English club manager m played for	143,064	String
Mannumengteams*	Number of English clubs manager m played for	154,434	Count
Manenddate	Date of last game of manager m in sample	154,434	Continuous
Manage	Age of manager m in days	152,429	Continuous
Manobs	Total game observation with manager m	154,434	Continuous
Manid	Identifier manager m counting spells after foreign employment as new manager.	154,434	Count
Manisexp	Manager m experience in sample, expressed in games	154,434	Continuous
Manperiodisexp	Manager m experience in sample during current uninterrupted employment in England, in games	154,434	Continuous

Manperiod	Counts the uninterrupted periods of employment in England	154,434	Count
Manexp	Cumulative experience manager m up to game g	154,434	Continuous
Manengexp	Cumulative experience manager m in England up to game g	154,434	Continuous
Manengexp	Cumulative experience manager m in UK and Ireland up to game g	154,434	Continuous
Mantenure	Manager m tenure in current spell	154,434	Continuous
Spellmanendtenure	Manager m tenure at end of current spell	154,434	Continuous
Spellid	Identifier for employment spell of manager m at club c	154,434	Count
Spellstartdate*	Startdate of employment spell in stata format	154,434	Continuous
Spellenddate*	Enddate of employment spell in stata format	154,434	Continuous
Spellendactive*	Employment spell is active at end of sample period = 1, 0 otherwise	154,434	Indicator
Spellstartactive*	Employment spell is active at start of sample period = 1, 0 otherwise	154,434	Indicator
Spellstartyear*	Season t in which current employment spell starts	154,434	Count
Spellendyear*	Season t in which current employment spell ends	154,434	Count
Spellstartperiod*	Period in which current employment spell starts	154,434	Count
Spellendperiod*	Period in which current employment spell ends	154,434	Count
Spellmanstartexp*	Total experience of manager m at start of employment at club c expressed in games	154,434	Continuous
Spellmanstartengexp*	Experience in England of manager m at start of employment at club c, expressed in games	154,434	Continuous
Spellmanstartukeireexp*	Experience in UK + Ireland of manager m at start of employment at club c, expressed in games	154,434	Continuous
Spellmanstartage*	Age of manager m at start of employment at club c, expressed in days	152,429	Continuous
Spellmanendexp*	Total experience of manager m at end of employment at club c, expressed in games	154,434	Continuous
Spellmanendengexp*	Experience in England of manager m at end of employment at club c, expressed in games	154,434	Continuous
Spellmanendukeireexp*	Experience in UK + Ireland of manager m at end of employment at club c, expressed in games	154,434	Continuous
Spellmanendage*	Age of manager m at end of employment at club c, expressed in days	152,429	Continuous
Spellstartdiv*	Division club c plays in at start of manager m employment spell	154,434	Count
Spellenddiv*	Division club c plays in at end of manager m employment spell	154,434	Count
Spellmanentry*	Manager m's first employment = 1, 0 otherwise	154,434	Indicator

Spellmanengentry*	Manager m's first employment in England = 1, 0 otherwise	154,434	Indicator
Spellmanukeireentry*	Manager m's first employment in UK + Ireland = 1, 0 otherwise	154,434	Indicator
Mannumspells*	Total number of employments spells manager m has in sample.	154,434	Count
Spellmanfinal*	Manager m's last spell in sample = 1, 0 otherwise	154,434	Indicator
Nextspelldiv*	Division in which manager m starts next employment. Lower than 4 th tier = 5, foreign = 2, big5 =1.	144,956	Count
Mancareerprogress*	Manager m's career progress after end of current employment. 0 = end career, 1 = new spell in lower tier, 2 = new spell in same tier, 3 = new spell in higher tier.	144,928	Count
Manentryperiod*	Period of manager m's first employment spell	154,434	Count
Manentrydiv*	Division of manager m's first employment spell	154,434	
Manengentryperiod*	Period of manager m's first employment spell in England	154,434	
Manengentrydiv*	Division of manager m's first employment spell in England	154,434	
Manendperiod*	Manager m's last active period in data	154,434	

2. Do-files for analysis

2.1. Step 1: estimating worker ability

It is important to note that running this step of the analysis on a desktop or laptop PC can take a significant amount of CPU time, i.e. days up to weeks depending on the machine. To alleviate this problem the code can be run in parallel on several processors or several machines. Each do-file allows to specify a start and end period at the top of the file. The file will run the analysis for all periods between the defined start and end period and store results in files denoted by the number of the end period. The suffix *'maxperiod'* in the name of the dta-files takes the number of the estimation run's final period. Keep in mind that the earlier periods typically run much quicker as they require smaller matrix operations in the estimation of the worker effects. For example, it takes about the same time to run periods 100-250 as 375-380 on the corresponding author's machine. The last code file can be adapted to merge the analysis runs back together. Currently, the runs are defined as 100-250, 251-300, 301-330, 331-355, 356-370, 371-376, and 377-380. We recommend creating a separate do-file for each run. We provide the resulting (intermediate) dta-files of this entire step, merged over the sample periods, to allow replicating the results without going through step 1. The corresponding do-files are:

- **Sumstatsteams.do**: Creates summary statistics at employer level. These are reported in **Table 2**.
- **movingestimationAKM.do**: File to estimate all variants of the worker effect regression (equation 5 in paper) as reported in Table 3 and Table B1-B2. The code stores results in the following files:
 - `estimatesperioddivision`maxperiod'.dta`: intermediate data file saves the estimated coefficients of covariates in regression of equation (5) without the manager and club effects by estimation run (i.e., minperiod to maxperiod). Organized at the level of estimation period – division pairing.
 - `datamoving`maxperiod'.dta`: data file stores all estimated worker and club effects at game level. Becomes a very large file if number of periods in the run increases.
- **movingestimationAddWin.do**: File to estimate the add win percentage measure as described in equation 6 in paper.
 - `winpercmoving`maxperiod'.dta`: Data file stores the added win percentage measure for each manager at the game level.
- **entryanalysisdata.do**: File takes the previously created files and merges them together into a unified file for further analysis. It calculates the baseline performance measures such as win% and goals scored/conceded by manager and estimation period. The output files are:
 - `estimatedcoefficients.dta`: data file saves the estimated coefficients of covariates in regression of equation (5) without the manager and club effects for the full sample. Organized at the level of estimation period – division pairing. We do not report these estimation results in the paper.
 - `managerbasequality.dta`: File stores the baseline worker performance measures calculated in this do-file at the worker level. Each measure – estimation period combination leads to a separate variable, meaning there are 281 variables per measure. The variables are set to missing if the worker has not been active in the dataset until the period examined.
 - `managerquality.dta`: File stores all estimated worker ability measures at worker level. Each ability measure – estimation period combination leads to a separate variable, such that each ability measure has 281 variables, i.e., period 100-380. The variables are set to missing if the worker has not been active in the dataset until the period examined.
 - `teamquality.dta`: File stores all estimated club effects at club level. Each ability measure – estimation period combination leads to a separate variable, such that each ability measure has 281 variables, i.e., period 100-380. The variables are set to missing if the club has not been active in the dataset until the period examined.
 - `entrantquality.dta`: File stores measures to characterize the ability of recent entrants at the division – period level. These include the number of entrants (prefix: *count-*), average

ability (*av-*), deciles of ability (*dec1/9-*). These can be over 5 years (*-5y-*) or 10 (*-10y-*) year intervals and by division (*-d1/4-*).

- *dataentryanalysisfull.dta*: A merged intermediate file which contains all estimation results and calculated measures at game level. In this file the performance measures are attributed to the period of the game, such that they are reduced to one variable per performance measure.
- *dataentryanalysiscollapsed.dta*: File at the employment spell level. All variables are attributed to start (*start-*) or end (*end-*) of the spell. This collapse is necessary to facilitate further analyses. This file is the main input for the do-file in step 3.
- *Managercareerprofiles.dta*: File at the worker-period level which covers the worker ability estimates combined with a set of variables from the source file. It allows to track the evolution of the worker's estimated ability over the entire career path. This file is the main input for the do-file in step 2.

2.2. Step 2: Establishing hiring thresholds and career analysis

This step of the analysis corresponds to the do-file **careeranalysis.do**. It creates a number of figures and tables which appear in the final paper. In addition, it calculates the values for the 'mediocre' threshold for different estimators of worker ability. Its input file is *managercareerprofiles.dta*, meaning it operates at the worker-period level of observation. The code is separated in six parts.

- **Part 1**: Calculates the 'mediocre' thresholds for every ability estimator from the do-file *movingestimationAKM*. We start from the average entrant ability (substandard threshold) and then increase this threshold in steps. We calculate the active ability improvement for each step drawing replacements from the entrant population. We figure out which step leads to the largest improvement and select this one. We do this 200 times to deal with the random component of the drawing of entrants. We select the step which is the mode of the maximum ability over the 200 runs as the final threshold for each estimator. Line 314 calls these modes, which we need in step 3. For example, for the baseline measure we find a threshold of mean entrant + 0.23. We store the results in the file *thresholdsearch.dta*.
- **Part 2**: For the baseline measure we calculate the ability improvement in the active workforce and bootstrap the entire selection procedure. This allows to generate the results in **Table 5** (line 479) and **Figure 4** (line 489 and 494).
- **Part 3**: Repeats the calculation of the mediocre threshold for the ability measures from the do-file *movingestimationAddWin*. The threshold is given on line 639. We store the simulation results in *thresholdsearchaddwp.dta*.

- **Part 4:** This sketches the illustration graphs of estimated ability over a worker career in **Figure 2** in the paper.
- **Part 5:** This makes the summary statistics of the personal characteristics and ability estimates at the worker level reported in **Table 1**, **Table 3**, and **Table B1**.
- **Part 6:** Performs the supporting analyses reported in Appendix D. The output regression tables are stored in **Table D1** (reported results) and **TableD1 extra** (non-reported results). We store the intermediate results on the convergence of worker ability estimates in learning.dta. We use this file to draw **Figure D1**, which we also repeat for the addwin% measure in the do-file (not reported).

2.3. Step 3: Analysis of hiring decisions

This step of the analysis corresponds to the do-file entryanalysis.do. It creates a number of figures and tables which appear in the final paper. Its input file is dataentryanalysiscollapsed.dta, meaning it operates at the employment spell level of observation. The code is separated in six parts.

- **Part 1:** Makes the correlation matrix of the varying ability estimates as reported in **Table B2**.
- **Part 2:** Generates a couple of intermediate variables needed in the analysis by merging in the file financialdataentry.dta, a file containing 3 additional variables brought in for the revision of the paper. These variables are the wage/turnover ration (wagerev) and the relative wage and revenues of the club, i.e. wage/annual mean wage and rev/annual mean rev.
- **Part 3:** Calculates how many spells start with a mediocre/substandard hire and how many workers are rehired as mediocre/substandard at least once in their career. This output is given in **Figure 3**, **Table 4**, and **Table B3** in the paper. The output for the tables appears from line 451.
- **Part 4:** Establishes how many mediocre/substandard hires improve over the employment spell to above the relevant thresholds. This output is summarized in **Table 7**.
- **Part 5:** Alternative analyses to confirm the results in Table 4, not reported in final paper, among others looking at the deciles of the entrant distribution.
- **Part 6:** Models the choice of an employer as in regression equation (7) in the paper and looks at career dynamics as reported in appendix D. The corresponding results appear in **Table 6** of the paper and **Table D2**. A set of additional (unreported) results appears in **TableD2 extra**.

