First-Author Determinants: An Empirical Analysis

C. Mirjam van Praag* , Bernard M.S. van Praag

Current and First Draft: March 13, 2003

ABSTRACT

This paper reports the results of an empirical analysis of name ordering strategies used for multi-authored academic papers in economics. We distinguish two name –ordering strategies: alphabetic and non –alphabetic. We investigate two questions. (1) What are the determinants of an author group's name ordering strategy? And (2) Is scientific standing or output affected by the relative alphabetic position of one's last name, given the dominant usage of alphabetic name ordering? We find that the distribution over the author group of relative costs and benefits of being a first author is clearly accounted for when positioning authors of a group. Economists are correct in perceiving name ordering as a deliberate decision. The usefulness of this deliberation is supported by the answer to our second question: is scientific standing lower for authors whose names rank low in the alphabet? Career prospects are better for academic economists who have high chances to be a first author, though this effect does not become clear instantaneously, but when an economist's career is more advanced and reputation and visibility might have been achieved already.

1. Introduction

The measurement of academic output has become a profession by itself. By and large, academics agree that research output in terms of publications is the major basis for performance measurement, and therefore for salary increases, promotions, outside offers, and last but not least reputation (Moore et al, 2001). This explains the growing presence of counting systems based on some explicit factors such as length, page size and scientific weights of journals. The scientific weight of a journal is nowadays typically based on the number of citations of articles that have appeared in the journal over the last couple of years and is called impact score.

Another explicit factor that significantly affects individual output of an academic is the number of co-authors that is identified on the paper. Most of the "accounting systems" divide research output by a factor (n+1), where n is the number of authors. This factor already indicates that individual output is increased by cooperation, given identical contributions. Indeed, the occurrence of multi-authored papers has increased dramatically over the last few decades, although this development might also be due to ongoing specialization and the development in communication technology (Hudson, 1996). For instance, Hudson (1996) counted that the occurrence of multi-

* Corresponding author. The authors are very grateful to Anne Marie LeClercq for her research assistance. Moreover, we express our gratitude to Martin Brands for his database assistance and for sharing his knowledge about how to retrieve the required information from our digital library. We

authored papers in two leading economics journals, i.e., *Journal of Political Economy* and the *American Economic Review*, have increased from six (JPE) and eight (AER) percent in 1950 to 40% and 55% in 1993.

An implicit factor that is widely believed to affect individual academic performance is the name order amongst the authorship. Name order selection for multi-authored papers in leading economics journals and its effects on individual productivity is the topic of our study. Engers et al. (1999) provide a theoretical explanation for the persistent use of alphabetic name orderings on academic papers in economics: Approximately 85% of multi-authored economics papers uses an alphabetic name ordering.

"In a context in which market participants are interested in evaluating the relative individual contribution of authors, it is an equilibrium for papers to use alphabetic name ordering. Moreover, it is never an equilibrium for authors always to be listed in order of relative contribution. In fact, . . . the alphabetic name ordering norm may be the unique equilibrium… Finally, we characterize the welfare properties of the non-cooperative equilibrium and show it to produce research of lower quality than is optimal and than would be achieved if co-authors were forced to use name ordering to signal relative contribution" (cited from Engers et al., 1999)

The occurrence of alphabetic name ordering in 85% of multi-authored papers implies that 15% of multi-authored papers in the economics literature use a different strategy of name ordering, which will in all cases be classified as "contribution based". As we will quantify by means of our very simple statistical framework, this percentage is much larger in reality, since some of the apparently but coincidentally alphabetic name orderings should also be contribution-based: Professor Aa can never signal his or her superior contribution by name ordering.

Our aim is to assess empirically the determinants of "contribution based" name ordering, thereby diverging from the equilibrium alphabetic convention, as derived by Engers et al. We shall also empirically investigate whether the use of alphabetic name order negatively affects the scientific output of economists who happen to have a name starting with a "Z" or the like. If that would be the case, we would establish evidence that name ordering is important indeed.

---

thank George Baker and Hessel Oosterbeek for their comments. The name order strategy of the authors is "non-alphabetic".

The paper proceeds as follows. Section 2 describes the simple statistical setup and thereby leads to a quantification of the real percentage of published papers with diverging name ordering strategies. Section 3 introduces the research questions. Section 4 describes the data we have gathered. Section 5a discusses the research results that lead to the empirical determinants of following an alternative strategy. Section 5b discusses the determinants of scientific production and the (perceived) drawbacks of having a last name late in the alphabet, given the alphabetic convention. Section 6 concludes.

2. Set up

Basically two name ordering strategies are used. The first is the alphabetic strategy and we call the second the alternative strategy, which is typically contribution- or merit based. We denote the alphabetic strategy with α, the alternative, (often) merit-based strategy with $\bar{\alpha}$.

Looking at a sample of two-authored articles by A and B, it is obvious that if all authors follow α, we would find that the fraction of AB articles is 100%. In the other extreme case, in which all authors would follow the merit strategy, we should find a fraction P(AB) of 50%, assuming independence between alphabetical order and merit. In practice part of the AB- papers is the product of the merit-strategy as A made the most important contribution indeed, while the other part is a result of the alphabetic strategy. In the case of two authors, the chance of finding an alphabetical ordering is

$$P(AB) = p_\alpha + (1 - p_\alpha) * 0.5 \qquad \rightarrow p\alpha = 2 * P(AB) - 1.$$

$$\rightarrow P(BA) = (1 - p_\alpha) * 0.5$$

Where p_α denotes the frequency of using an alphabetic strategy. The chance P(AB) is the sum of the probability of using the alphabetic strategy and the probability of using the alternative strategy times half. It seems that the estimation of p_α is interesting *per se.* For instance, an α - frequency of 0.8 would indicate that an alphabetic strategy is used in 60% of cases, while in the other 20% of the alphabetic cases the order reflects the relative contribution.

For three authors A, B, and C, we find similarly:

$$P(ABC) = p\alpha + (1 - p\alpha)*1/6 \rightarrow p_3 = 6/5 * P(ABC) - 1/5$$

as pure randomness would give 3! = 6 possible combinations. Likewise, we find for *n* authors the following relationship between the observed frequency of alphabetic name ordering and the frequency of following an alphabetic strategy:

$$P(AB\,..\,N) = p_\alpha + (1 - p_\alpha) * 1/n! \rightarrow p_\alpha = \frac{P(AB...N) - \frac{1}{n!}}{1 - \frac{1}{n!}} \qquad (1)$$

The larger n, the lower the difference between the observed alphabetic fraction and the fraction of users of an alphabetic strategy. It follows that rather rapidly with *n* increasing, we may identify the alphabetic order as giving no indication on relative contributions.

If we assume that $p_\alpha = p_\alpha(x,n)$, we find that

$$\frac{\partial p_\alpha}{\partial x} = \frac{\partial P}{\partial x}\frac{1}{(1 - \frac{1}{n!})}$$

while

$$p_\alpha(n) - p_\alpha(n-1) = \frac{P(AB...N;n) - \frac{1}{n!}}{1 - \frac{1}{n!}} - \frac{P(AB...N;n-1) - \frac{1}{(n-1)!}}{1 - \frac{1}{(n-1)!}}$$

As we shall estimate *P* but like to draw conclusions on p_α these relationships will be relevant.

3. Questions

There are several questions we would wish to answer by means of our empirical analyses. The first series of questions concerns the determinants of p_α:

1. Is p_α constant in n, i.e., the number of authors of a multi-authored paper?

Due to conflict avoidance it seems likely that it *increases* with the number of authors. However, the more authors there are, the more likely it is that individual contributions are not spread evenly. This would suggest that p_α *decreases* as *n* increases. It might also be the case that the likelihood of having an opponent to the alphabetic strategy amongst the authorship increases as the number of authors increase: this would also lead to a *lower* p_α. For instance, if age or gender differences within an author group would lead to deviations from an alphabetic strategy, it is more likely that this will be the case with more than with less authors. We answer this question by including the number of authors as an explanatory variable in the regression equation that explains

the name ordering strategy, but acknowledge that the divergence between *P(AB)*, what we observe, and p_α sharply decreases as a function of *n*. We shall therefore also test this hypothesis based on the descriptive statistics as presented in the next section.

2. To what extent does the distribution among the authorship of "scientific weight" and "scientific age" affect name ordering strategies?

It would be conceivable that more senior economists and/or economists with more publications care less about their future careers and make (the first) place for the younger economists. On the other hand, it might as well be conceivable that the more senior economist has contributed more than proportionally to the paper, which gives incentives for the older person to fight positions. These arguments would lead to the hypothesis that the larger the variance of seniority is within the group of authors, the smaller the fraction p_α will be.

If the complete authorship is on average further in their careers, they care less and avoid conflicts by choosing the (default) alphabetic name ordering strategy. However, if all co-authors would be young and eager they would have more incentive to struggle and therefore follow a merit strategy.

We test these hypothesized distributions of incentives by including the averages and variances of "scientific weight" and "scientific age" as dependent variables. We measure "scientific age" as the years passed since an author's first publication. We measure "scientific weight" as an author's total number of publications.

3. Is the fraction p_α determined by the relative order in the alphabet of the last names of the authorship?

One might expect that a Z-author will accept the alphabetical strategy less easily than an A-author, because a Z-author will practically never have a chance of becoming the first author. An A-author in combination with a B-author and a C-author will not think it worth the effort and conflict to deviate from the default alphabetic strategy: they will all get their turn next time. However, this might be very different for an XYZ combination. We test this hypothesized relationship between the fraction p_α and alphabetic position by including a regressor that denotes the average alphabetic position amongst the authors. We measure an individual's alphabetic position by translating an A into a 1 and a Z into a 26. Alternatively we transform the alphabetic order logarithmically from 0 to ln(26).

It is furthermore conceivable that a larger spread in the distribution of names over the alphabet will influence the adherence to an alphabetic name ordering strategy. We therefore also include the variance of the alphabetic positions of the authors as a potential determinant of the fraction p_α.

4. Does the convention of using an alphabetic strategy depend on the geographical location of the authors?

We shall assess whether combinations of authors affiliated with an institute in the US use strategies different from European groups or Asian groups or combined groups.

5. Is the fraction p_α decreased in case of the presence of a female co-author?

It might be the case that the rare females in our profession are treated with courtesy in this respect.

6. Does the fraction p_α decrease when there is a non-academic co-author?

We think that authors outside academia simultaneously care less about their academic proceedings and will often be the less contributing partner in the publication venture. This might lead to a decreased p_α.

7. Is the name ordering strategy affected by the extent to which a publication counts for someone's academic merit?

We test whether the number of pages of an article as well as the impact factor of a journal affects p_α. The expectation is that the more important a publication is for someone's output score, the higher the stakes, and the higher the incentive is to stick to the default alphabetic name order in order to avoid conflicts of a larger size.

8. Are p_α and its determinants different for long term versus occasional combinations of authors?

We would think, based on our own experience, that occasional groups of authors would rather follow a default strategy, whereas longer-term relations would alternate their names and therefore easier deviate from the alphabetic order. Conflicts can be solved more easily in a repeated game, by making promises (or threats) about name order in the next period.

The second series of questions relates to individual authors: is an individual's scientific output affected by his/her last name? Is it easier for A-authors to achieve success in our business than for a Z-author, as the latter will more often be the last author and therefore obtain less credits, visibility and/or citations? Or alternatively, is

it easier to become successful as a Z-author than as an A-author since the latter will never have the opportunity to signal a more than proportional contribution to a publication, given the alphabetic convention? We test this type of questions by estimating two performance measures and include the relative alphabetic rank of an author as a potential determinant. The two performance measures that we use are: (1) an author's total number of publications, i.e. "scientific weight" and (2) an author's scientific output per year, i.e. "scientific weight" divided by "scientific age". We estimate the relationship between these two performance measures and the alphabetic position of an author's last name, while controlling for gender, geographical location and whether the author works in or outside the university. We implicitly assume for all analyses that authors don't seek co-authors based on their name (occurring later in the alphabet than their own name). When answering the second type of questions, whether names affect scientific weight, we implicitly assume that all authors deviate to the same extent from alphabetic name ordering.

4. Data

The sample consists of all regular journal articles published in the period 1997-1999 in the following 11 journals:

- American Economic Review;
- Economica;
- Economic Journal;
- European Economic Review;
- International Economic Review;
- Journal of Economic Behavior and Organization.
- Journal of Economic Perspectives;
- Journal of Economic Theory;
- Journal of Political Economy;
- Quarterly Journal of Economics;
- Review of Economic Studies;

Book reviews and the like are excluded from the sample. We have selected these eleven journals for their general character, their rather high impact and their mix between European and American origin. Moreover, these are journals for which we know that the editors do not impose a specific (alphabetic) author ordering. We have

selected three quite recent years (at the time we started the fieldwork). We needed three years to obtain a large enough sample of multi-authored articles. The period selection is based on the logic that a recent selection of articles guarantees that the time dependent variables of our choice had approximately the same individual values in this period of publication as they had in the period that we gathered them. An example is an individual's score on the variable "scientific weight", i.e. his/her publication track record so far.

The resulting sample consists of a total of 2311 articles. These articles have been selected from the digital database "Web of Science". This resulted in our own (Filemaker) *articles* database. For each of the multi-authored papers, we have downloaded the following variables from the *Web of Science* into our articles database.

- Title
- Author names
- Number of authors
- Journal name and its impact score[1]
- Year and journal issue
- Number of pages
- Name ordering (alphabetic or not)
- Number of previous joint publications

Table 1 shows the numerical relationship between articles and authors. Conform Hudson's findings, 55% of these 2311 are multi-authored: we have 1,278 multi-authored observations in our database. The vast majority of multi-authored papers, i.e. three quarter, has been written by two authors. The average number of authors per article in our sample is 1.72.

Table 1 moreover shows the name ordering distribution for papers with more than one author. The observed percentage of alphabetic name ordering in the entire set of multi-authored papers is 88%, translating, by means of equation (1), into a fraction

[1] The impact factor is based on the objective ranking, the Social Science Citation Index (SSCI), annually published by the Institute for Scientific Information (ISI). This index gives for thousands of journals the number of citations each year. This leads to an impact factor for each journal, using the following formula: impact factor year X = (Cites in year X to articles published in year X-1 and in year X-2)/(Number of articles published in year X-1 and in year X-2). We use X=1999.

that uses an alphabetic strategy, p_α, of 80%. This overall p_α is calculated as the weighted average of the p_α's for each n. The observed alphabetic percentage is higher the fewer authors are involved: it is 91% for two authors and 57% for five authors. As equation (1) already showed, the divergence between the observed alphabetic fraction and the fraction of users of an alphabetic strategy decreases as n increases: the difference has totally vanished with n=5 already. We still observe that p_α decreases as n increases: two authors use an alphabetic strategy in 81% of cases, whereas this percentage decreases to 79% for three authors, 67% for four authors, to 57% for five authors.

We conclude that the alphabetic strategy is used less whenever the number of authors increases. This might be due to the fact that the probability of an uneven spread of real contributions amongst authors increases as the number of authors increases.

The next step was to build a (linked) *authors* database. The unit of observation in this second database is not articles but authors instead. The authors' dataset contains the following author specific variables for each of the 2103 different authors:

Table 1

Frequency distribution of number of authors and name ordering over articles

Number of authors	Number of articles	%	Number of authors involved	Number alphabetic name ordering	P(AB) (%)	Pα (%)
1	1033	45	1033			
2	946	41	1892	858	91	81
3	282	12	846	232	82	79
4	41	2	164	28	68	67
5	7	0	35	4	57	57
6	1	0	6			
8	1	0	8			
>1	1278		2921	1122	88	80
Total	2311	100	Different authors: 2103			

- Name
- Gender
- Geographical location of main activities (US/Europe/Asia/Other)
- Type of institution of main base (university or not)
- "Scientific age" (number of years since first publication in any journal with impact factor)

- "Scientific weight" (Number of publications in all journals with an impact factor since 1969)

These variables are retrieved from *Econlit*, which includes publications beginning in the year 1969.[2] It also includes first names of authors, which turns out useful for determining an author's gender.[3] Gender is known for 98% of authors. Table 2 shows that 89% of them are males.

Table 2

Distribution of authors sample: gender, institution, geographical location.

	Total		University		US		Europe		Asia	
	Freq.	Perc.	Freq.	perc.	Freq.	Perc.	freq.	Perc.	Freq.	Perc.
Male	1861	88.5	1681	90.3	1049	56.4	617	33.1	87	4.7
Female	242	11.5	211	87.2	145	59.9	80	33.1	6	2.5
Total	2103	100	1892	90.0	1194	56.8	697	33.1	93	4.4

The type of institute each author works for, i.e. university or not, is retrieved from *Econlit*. The result is shown in Table 2: 90% of the authors is mainly affiliated with a university.[4] For the geographical location of the institute that an author is affiliated with, we distinguish between the US, Europe, Asia and other. Table 2 shows that 57% of the authors is US-based, 33% Europe based and only 4% Asia-based. The remaining 6% works elsewhere.[5] These patterns do not significantly differ over the sexes.

[2] We didn't use the *Web of Science* database as we did for the articles related variables because it includes publications as of 1988 only, whereas *Econlit* includes articles beginning in the year 1969. This is clearly preferable for obtaining as much as possible unbiased and untruncated values for "scientific weight" and "scientific age".

[3] For 196 authors whom we didn't know personally and whose first names were not clearly identifiable as male or female (we had for instance difficulties with many of the Chinese and Japanese names or with Western names that are used for both males and females) we had to use an additional procedure. We first looked up their personal websites through the search engine Google. These often give a clue about gender, but even when a picture was shown, we could sometimes not draw a definite conclusion. We gathered the email addresses of authors without identifiable personal website (or whose pictures are not decisive on the issue at hand) and sent them a mail with our information request. We obtained many responses and we were finally left with 45 authors whose sex we still don't know. We dropped them from the sample for regression analyses.

[4] If an author has changed affiliations over time (to or from university from or to other), we have counted the most recent affiliation, unless the author has worked in a different institution for the majority of the time period in which most of his/her publications within our database have been produced. Whenever an author works for both types of institutions, we consider the institution that has been mentioned first (in the author affiliations footnote in the article) as his/her main affiliation.

[5] If an author has worked in several institutes over the relevant period and if these institutes are located in different geographical areas according to our definition, we have assigned both values to the geographical location variable.

To calculate an author's scientific age, i.e. the number of years (s)he is "in the business", we used *Econlit* to find his first publication in a journal with a positive impact score. Scientific age is defined as 2002-the year in which the first article was published in any journal. Unfortunately, the "scientific age" variable is truncated at 33 since we were unable to trace back any publications before 1969. Figure 1A shows the (cumulative) distribution of scientific age. The median value is 12.8 years, and the mean value is 14.7 years. The truncation at 33 years of scientific age seems to be not too severe a problem: the scientific weight of a maximum of 3.4% (72 authors) is truncated.

The variable "scientific weight" denotes the number of articles the author has published in any economics journal with positive impact according to Econlit and therefore only as of 1969. The source Econlit, however, generates an additional bias since it spells out the names of at maximum three authors: It mentions the name of the first author only whenever an article has been written by more than three authors. Consequently, an author's track record according to Econlit is the sum of the number of articles (s)he has written with two co-authors at a maximum and the number of articles (s)he has written as first co-author with more than two co-authors.

To give an indication of the scale of the problem in our sample, we calculate that with less than 4 percent of the articles, the scientific weight of 5.5 percent of the authors, i.e. the number of co-authors minus one times the number of articles written by more than three authors, is negatively affected by this omission. The majority of these authors will have last names starting with letters late in the alphabet, given the alphabetic convention in our profession. To cope with the problem, we have omitted any article with more than 4 authors from the "scientific weight" calculation for their first author.

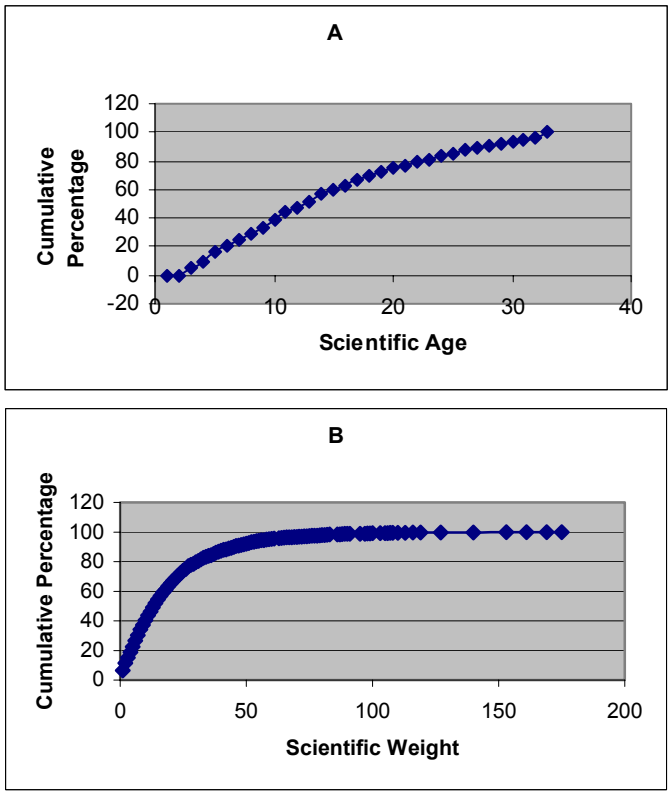
However, since our own database is retrieved from *WOS*, it does not suffer from this problem. This would lead to an upward bias in the "scientific weight" of authors who have contributed to articles with more than 4 authors that happen to belong to our database. Therefore, we have also omitted the articles written by more than 3 authors available in our database from the calculation of the "scientific weight" variable. The result is that there is no bias anymore.

Figure 1B shows that scientific weights vary from 1 to 175. The median value is 13.5 articles, whereas the mean is 20. The correlation between scientific weight and age is a positive 66.8%, as might have been expected.

5a. Determinants of non-alphabetic strategies

A probit regression is executed to test the first set of hypotheses that are all related to the determinants of alphabetic name ordering strategies. The dependent dummy variable "name order" in this regression equation equals 1 when a group of authors deviates from the conventional alphabetic name ordering strategy, whereas it is zero when they order their names alphabetically. We only observe the resulting name ordering of author groups, and unfortunately and obviously not the strategies they follow. The only certain correspondence between the observed name order and the strategies followed is the observation of a non-alphabetic name order, which clearly corresponds to the use of a deviating strategy.

Figure 1 Cumulative distribution of authors' scientific ages and weights



The independent variables used are:

- "*authors*": the number of authors to answer the first question whether more/less authors lead to (non)alphabetic name ordering
- *"avlogweight*", the average (logarithm of the) scientific weight of the authorship and "*avlogage*", the average (logarithm of the) scientific age of the authorship to test whether more advanced authorships care less about their career advancements and therefore use the default (alphabetic) name ordering strategy. To complete the test of our hypotheses concerning the distribution of scientific standing amongst the authorship as described in question 2, we also include the standard deviation of scientific weights ("*sdlogweight*") and the standard deviation of scientific ages ("*sdlogage*") into the regressions. We expect an increasing variance of standing amongst the authorship to decrease the usage of the alphabetic strategy (see question 2). Since the correlation between scientific weight and age is fairly high (67%) we also estimate relationships in which the distribution of either age (avlogage and sdlogage) or weight (avlogweight and sdlogweigth) is included.
- In order to answer the third question, whether the distribution of the last names of the group of authors affects the name ordering strategy, we include the average position of the names in the alphabet, "*avalphabeticpos*" (where an A gets value 1 and a Z value 26) as well as the variance of relative name positions in the alphabet, "*sdalphabeticpos*". We expect that author groups with on average names further in the alphabet will be inclined to deviate from the alphabetic strategy, as will author groups with a larger spread over the alphabet in their names.
- To answer the questions four to six about geographical location, presence of a female amongst the authors, and presence of co-authors with a non-academic affiliation respectively, we include (i) the dummy variable "*all-us*" that takes on value 1 for groups of US authors exclusively, (ii) the dummy variable "*female*" that is one whenever at least one of the co-authors is female, (iii) the dummy variable "*non-academic*" which is one when at least one author has a non-academic affiliation.
- The seventh question whether name ordering practices depend on the academic importance of a publication, can be answered by including a variable "*impact*" that reflects the impact score of the journal and a variable "*pages*" that reflects the number of pages of the article the group has written.

- The last question, dealing with the possibility that author groups have a different strategy for a 'one-shot' than for a long –standing relationship, is answered by including a dummy variable "*oneshot*" which is one for groups that have not previously cooperated (insofar as this has led to a publication) and zero otherwise.

Table 3 shows the estimation results. As was shown by the descriptive statistics, larger author groups are less inclined to stick to the conventional alphabetic name ordering strategy. The estimated effect, as shown by the elasticity (taken at the average value of the independent variables), is a 5% increase in the observed fraction of non-alphabetic name ordering per additional author. As was previously pointed out, this effect comprises both the decreasing divergence between what we observe and what strategy is used when the number of authors increases as well as the decreased usage of alphabetic name ordering strategies. This answers the first question.

Our expectations about the effect of the distribution of academic standing amongst the author group (question 2) are confirmed by the results: author groups with a higher average level of scientific age or weight tend to stick to alphabetic name ordering. The spread of scientific weight or age amongst the author group affects the use of alphabetic name ordering also: the higher the spread, the higher the probability of non-alphabetic name ordering. This can be explained by higher probabilities of uneven contributions and/or by higher variances in the individual benefits of being a first author. The effects of the distribution of scientific weight and age are in the same direction, though the first column of Table 3 shows that scientific weight overrules the effect of scientific age.

The expected answer to the third question of this paper is also supported by the findings: XYZ-authors are more inclined to deviate from alphabetic name ordering than are ABC authors, probably due to their expectation that they will not get another chance to be a first author next time. Moreover, the data confirm that a larger spread in the distribution of names over the alphabet increases the adherence to an alphabetic name ordering strategy: A and Z will deviate less easily than will A and B or Y and Z.

The effects of geographical location, presence of a female, or a non-academic co-author are all insignificant from zero. We had no particular upfront expectations about geographical location, but we thought the presence of females and people with a non-academic affiliation in the author group to decrease the probability of alphabetic name ordering.

The impact figure of a journal has no effect on the usage of alphabetic name ordering, though the number of pages clearly has: the longer the article, and more powerful the effect of a publication for the authors' careers, the higher the probability that the default alphabetic name ordering strategy is followed. This confirms the expected answer to question seven in line with conflict avoidance.

Table 3 Determinants of non-alphabetic name ordering

	Probit 1			*Probit II*			*Probit III*		
Variable	DF/dx		*s.e.*	DF/dx		*s.e.*	DF/dx		*s.e.*
Authors	.05077	***	.0152	.05002	***	.0151	.05712	***	.0155
Av-logweight	-.05047	***	.0172	-.05445	***	.0121			
Sd-logweight	.04242	***	.0134	.03901	***	.0097			
Av-logage	-.01035		.0317				-.08382	***	.0226
Sd-logage	-.01905		.0521				.10485	***	.0377
Av-alphabetpos	.00363	**	.0017	.00360	**	.0017	.00356	**	.0017
Sd-alphabetpos	-.00063	**	.0003	-.00063	**	.0003	-.00055	*	.0003
All-Us[d]	.00158		.0179	.00104		.0178	.00819		.0182
Female[d]	.00644		.0213	.00698		.0213	.01874		.0226
Non-academic[d]	-.01888		.0217	-.01847		.0217	-.01204		.0230
Impact	.00613		.0105	.00627		.0105	.00632		.0107
Pages	-.00288	***	.0008	-.00286	***	.0008	-.00305	***	.0009
Oneshot[d]	.01734		.0189	.01672		.0188	.04546	**	.0185
Pseudo R2	0.120			0.119			0.096		
Loglikelihood	-393			-393			-404		
Obs. fraction non-alphabetic	0.118			0.118			0.118		
Pred. fr. non-alph. (at av x)	0.092			0.092			0.096		
N	1233			1233			1233		

*10% significant; **5% significant; ***1% significant; ([d]) dF/dx is for discrete change of dummy variable from 0 to 1

The table finally shows that occasional author groups behave, on average, no different than do longer term author teams. Table 4 shows what name order strategies repeated author groups (at least one of the articles they have written turned up in our database) have used for their joint articles that appeared in journals (beginning in1969). It illustrates the surprising finding that one shot and repeated author groups do not differ in their name order usage: the percentage of "always alphabetic" is not significantly different from the average of 88% as shown in Table 1. The table

moreover shows that almost 5% of the non-occasional teams uses an alternating strategy.[6]

5b. Benefits of Being a First Author and Costs of last name "Z"

Outside economics merit strategies are quite commonly used. This will give non-economists an inclination to perceive the contributions of first authors in economics journals as the author with the highest contribution. Moreover, a common assumption amongst economists ourselves is that a deviation from alphabetic name ordering is a signal about contributions. Therefore, being a first author implies certain benefits.

Table 4 Longer term author teams and their name ordering strategies for their joint papers as of 1969

Strategy	Number of articles	Percentage
Always alphabetic	599	85.7
Always non-alphabetic	24	3.4
Alternating	33	4.7
Other	43	6.2
Total	699	100

However, this is certainly not all. A first authorship entails more benefits than the suggestion of being the major contributor. Citation indices, for instance, have for a long time only counted the names of first authors. This implies an additional benefit attached to being the first author, since the number of citations is a performance measure that is commonly used. Moreover citations within articles, which clearly contribute to someone's reputation and visibility, are abbreviated as "first author et al." as soon as there are more than two authors: all but the first author are no longer named in the main text of an article. Visibility is also constrained for others than first authors in frequently used search engines such as *Econlit*. As we encountered ourselves in our data gathering process, *Econlit* merely reveals the name of the first author for articles with more than three authors. In addition, all these benefits of being a first author might have been spin-offs to other benefits such as being invited to seminars and conferences (for keynote speeches), becoming a fellow in a high

[6] The alternating strategy is defined as follows: the first article an author group publishes will be in alphabetic name order. Subsequent to the first article, the second till nth articles will follow a different strategy, whereas the (n+1)th article is alphabetically ordered again. So out of any k articles by author groups of size n, we expect the percentage of alphabetically ordered articles to be 1/n for any k that is a multiple of n; 2/(n+1) for any k that is a multiple of n+1 et cet. We neglect the fact that the alternating author order will rather be manuscript based than publication based.

standing research school and so on and so forth. Since research time or input (and thereby output) within universities is allocated to individuals partly based on these fellowships and the like, we wish to test whether having a name ranked later (earlier) in the alphabet contributes to someone's academic standing. We control for other factors such as gender, country, and (other than) academic affiliation. We assume that all authors have deviated to the same extent from alphabetic name ordering.

Table 5 shows the estimation results where the indicator for academic standing is (I and II) log(scientific weight) in a linear regression model, (III and IV) log(scientific weight/ scientific age) in a linear regression model.

The first rows contain the key variables of interest. Based on the distribution of names over the alphabet, see Appendix Table 1, we have formed eight equally large groups based on name ranks: "nameAB" is the reference group that is one for authors whose names start with an A or B. The variable names in Table 5 are defined accordingly. We have moreover created a variable "letter", which is based on the cumulative distribution and ranges from 3.76/2 for authors whose names start with an A to (98.43+100)/2 for authors whose names start with a Z (see Appendix Table 1)[7]. These rows show that there is no significant effect at all of being an A-author or a Z-author, neither for scientific weight, nor for scientific weight per year.

The effects of the control variables are notable. Females perform less well, both in total and per year. A further consistent effect established is that economists from the US perform much better than their European colleagues, whereas the Asian economists perform poorer. (Non-)Academic affiliation has the expected effect: people with a university affiliation (possibly besides other affiliations) have a significantly higher scientific weight, whereas their colleagues with a non-academic affiliation (possibly besides an academic affiliation) have a (marginally significant) lower scientific weight.

Table 6 shows the results of analyses that were performed on a subset of the sample. It might be the case that being an A-author or a Z-author will only start to affect scientific production (for instance by an increasing ease of getting stuff published) after the first couple of publications (most likely as a first author for A-authors and as a last author for Z-authors). The logic of this hypothesis is as follows: at the start of careers, nobody has an advantage of reputation and visibility. However,

Table 5 Determinants of Scientific Weight

Variable	I OLS log weight (s.e.)	II OLS log weight (s.e.)	III OLS Log(weight/age) (s.e.)	IV OLS Log(weight/age) (s.e.)
Letter		.0002 (.0008)		.0003 (.0005)
Name CD	-.0686 (.0952)		.0101 (.0630)	
Name EFG	.0086 (.0949)		.0293 (.0627)	
Name HIJK	-.0293 (.0929)		.0081 (.0614)	
Name LM	.0583 (.0946)		.0528 (.0626)	
Name NOPQR	.0555 (.0925)		.0539 (.0612)	
Name S	.0127 (.0971)		.0286 (.0642)	
Name TZ	-.0634 (.0932)		.0160 (.0617)	
Female	-.8323 *** (.0746)	-.8273 *** (.0745)	-.4249 *** (.0494)	-.4233 *** (.0492)
US base	.1508 *** (.0491)	.1528 *** (.0489)	.0598 * (.0324)	.0597 * (.0323)
Asia base	-.4029 *** (.1280)	-.4169 *** (.1275)	-.2117 ** (.0847)	-.2163 *** (.0843)
Academic affiliation	.2814 * (.1566)	.2650 * (.1560)	.2408 *** (.1036)	.2370 ** (.1031)
Non-academic affiliation	-.2612 * (.1444)	-.2769 * (.1439)	-.0956 (.0955)	-.0994 (.0951)
Constant	2.298 *** (.1733)	2.300 *** (.1672)	-.3092 *** (.1146)	-.2871 ** (.1125)
N	2058	2058	2058	2058
Adj. R2	0.083	0.084	0.083	0.056

*10% significant; **5% significant; ***1% significant

after the first x publications, the A-author might have become more visible and of higher reputation than the Z-author which results in a significant effect of names on the cross-sectional variation of scientific weight amongst a group of relatively high weight authors. We test this hypothesis by performing the analyses for the group with a higher than median scientific weight only. Table 6 shows that a modestly strong effect of name rank in the alphabet pops up: *all* dummies are negative (as compared to the AB-group) and some of them are significant. "Letter" is also negative, but insignificant, probably due to the linear specification of a non-linear relationship. The effects of control variables are minimal within the subsample of authors of higher academic standing.

[7] The appendix table also shows that there is no significant difference over letter ranks in whether authors cooperate or prefer to publish alone.

Table 6 Determinants of Scientific Weight for academics with >median scientific weight

Variable	I OLS log weight (s.e.)	II OLS log weight (s.e.)	III OLS Log(weight/age) (s.e.)	IV OLS Log(weight/age) (s.e.)
Letter		-.0006 (.0005)		-.0006 (.0005)
Name CD	-.1332 ** (.0645)		-.0814 (.0596)	
Name EFG	-.0839 (.0646)		-.0828 (.0597)	
Name HIJK	-.1309 ** (.0632)		-.1010 * (.0585)	
Name LM	-.0699 (.0652)		-.0322 (.0603)	
Name NOPQR	-.0753 (.0623)		-.0884 (.0577)	
Name S	-.1245 * (.0664)		-.1549 ** (.0615)	
Name TZ	-.1178 * (.0639)		-.0526 (.0591)	
Female	-.2503 *** (.0781)	-.2427 *** (.0780)	-.4249 *** (.0494)	-.1052 (.0722)
US base	.0944 *** (.0334)	.0948 *** (.0333)	.0598 * (.0324)	.0047 (.0308)
Asia base	-.0668 (.1026)	-.0784 (.1024)	-.2117 ** (.0847)	-.0502 (.0948)
Academic affiliation	.0198 (.1162)	.0107 (.1158)	.2408 *** (.1036)	.0862 (.1071)
Non-academic affiliation	-.0846 (.1046)	-.0929 (.1039)	-.0956 (.0955)	.0418 (.0962)
Constant	3.469 *** (.1281)	3.418 *** (.1230)	-.3092 *** (.1146)	-.3557 *** (.1139)
N	952	952	952	952
Adj. R2	0.02	0.02	0.02	0.01

*10% significant; **5% significant; ***1% significant

6. Conclusion

Though the default name ordering for academic papers in the field of economics is alphabetic (Engers et al, 1999), we calculate that only 80% of multi-authored papers sticks to this alphabetic convention. The other 20% use a deviating (merit) strategy.

This paper's first aim was to evaluate whether there are systematic determinants of name order strategies amongst economists. The second objective of this paper was to establish whether academic standing is affected by having a last name that ranks low in the alphabetic distribution, given the default alphabetic strategy.

We establish our quite intuitively appealing results by taking a sample of multi-authored papers consisting of three complete years and eleven complete economics

journals. This resulted in a sample of 1230 multi-authored articles by 2100 economists based on which we perform various regression analyses.

The first set of analyses seeks the determinants of using a deviating name ordering strategy. We found that larger author groups deviate less easily from the alphabetic name ordering strategy[8]. Moreover, author groups with a higher average level of academic standing deviate less, since the relative benefits of visibility clearly decline over someone's career. Author groups with a higher spread in scientific weight or age appeared to deviate significantly more: uneven contributions and relative large benefits from being a first author are likely explanations for this. We have also looked at the effect of the alphabetic name ranking distribution (1-26) amongst the authorships. It was found that author groups with a higher average rank (i.e. YZ-groups as compared to AB-groups) tend to deviate from the alphabetic name order. AB-groups will get their turn next time, whereas YZ-groups won't. Author groups with a larger spread in the alphabetic ranks of their names tend to deviate less: a Z-author will more easily fight for a first position with a Y-author than with an A-author. We tried to establish whether the importance of the contribution of an article to someone's academic track record affects name ordering. This was measured by the inclusion of the impact factors of the journals in the sample and by the number of pages of each individual article. Impact has no impact, though the number of pages has: the longer an article is, the higher the likelihood of alphabetic name ordering. Conflict avoidance seems to be the explanation. We find no distinction in name ordering strategies between one-shot cooperative groups and repeated cooperatives.

The second set of analyses seeks to establish the individual determinants of academic standing. We find no effect at all of a person's name rank on his scientific output given the entire sample. However, it seems more likely that the effect of reputation and visibility will start working after a number of publications. Starters have all little visibility and reputation whether they are named Andrews or Zenner. We therefore also look at the effect of name ordering for the subsample of authors who might have gained a reputation already by their publications thus far. Indeed, we find (marginally) significant and quite consistent effects of the first letter of an economist's last name on scientific production given an author has already a certain standing and reputation in academia. Authors with names later in the alphabet have

indeed a slightly lower scientific production, both in total and per year than do AB-authors.

Other notable determinants of academic standing are gender, (non-)academic affiliation and whether one is based in the US, Asia or Europe.

One severe drawback is attached to our analyses: they will clearly suffer from a publication bias. Nevertheless, we can fairly well conclude from our first set of analyses that name ordering is no random process at all: economists clearly think it to be a strategic issue and strategies are adapted to certain characteristics of the author group. Our second set of analyses shows that academic economists are correct in their view that name ordering is a strategic decision: being a first author is beneficial for academic standing though not instantaneous but after a while.

References

Eisenhauer, J. G. 1997. "Multi-authored papers in economics." *Journal of Economic Perspecives*, 11:2, pp. 191-92.
Engers, M., J. S. Gans, S. Grant, and S. P. King. 1999. "First-author conditions." *Journal of Political Economy*, 107:4, pp. 859-83.
Hudson, J. 1996. "Trends in multi-authored papers in economics." *Journal of Economic Perspectives*, 10:3, pp. 153-58.
Moore, W. J., R. J. Newman, and G. K. Turnbull. 2001. "Reputational capital and academic pay." *Economic Inquiry*, 39:4, pp. 663-71.

[8] Even when you take into account that larger author groups that do not use an alphabetic name ordering strategy have a lower probability to still end up in alphabetical name order.

Appendix
Distribution of alphabetical ranking of author names

	Authors with co-authors			Authors without co-authors		
Ranking	Frequency	%	Cum. perc.	Frequency	%	Cum. perc.
1=A	79	3.76	3.76	25	3.81	3.81
2=B	188	8.94	12.70	49	7.46	11.27
3=C	144	6.85	19.54	51	7.76	19.03
4=D	105	4.99	24.54	28	4.26	23.29
5=E	55	2.62	27.15	16	2.44	25.73
6=F	74	3.52	30.67	24	3.65	29.38
7=G	125	5.94	36.61	29	4.41	33.79
8=H	114	5.42	42.04	44	6.70	40.49
9=I	23	1.09	43.13	3	0.46	40.95
10=J	41	1.95	45.08	16	2.44	43.39
11=K	100	4.76	49.83	28	4.26	47.65
12=L	103	4.90	54.73	34	5.18	52.83
13=M	153	7.28	62.01	52	7.91	60.74
14=N	46	2.19	64.19	18	2.74	63.48
15=O	36	1.71	65.91	9	1.37	64.85
16=P	92	4.37	70.28	37	5.63	70.48
17=Q	4	0.19	70.47	2	0.30	70.78
18=R	100	4.76	75.23	29	4.41	75.19
19=S	238	11.32	86.54	63	9.59	84.78
20=T	68	3.23	89.78	21	3.20	87.98
21=U	9	0.43	90.20	0	0.00	87.98
22=V	54	2.57	92.77	20	3.04	91.02
23=W	102	4.85	97.62	32	4.87	95.89
24=X	3	0.14	97.77	3	0.46	96.35
25=Y	14	0.67	98.43	11	1.67	98.02
26=Z	33	1.57	100	13	1.98	100
Total	2103	100		657	100	