DISCUSSION PAPER SERIES

IZA DP No. 905

Rhetoric in Economic Research: The Case of Gender Wage Differentials

Doris Weichselbaumer
Rudolf Winter-Ebmer

October 2003

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Rhetoric in Economic Research: The Case of Gender Wage Differentials

Doris Weichselbaumer
University of Linz

Rudolf Winter-Ebmer
University of Linz, Institute for Advanced Studies Vienna, CEPR and IZA Bonn

Discussion Paper No. 905
October 2003

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 905
October 2003

ABSTRACT

Rhetoric in Economic Research: The Case of Gender Wage Differentials*

Scientific rhetoric can have a profound impact on the perception of research; it can also drive and direct further research efforts. What determines whether results are discussed in a neutral or a judgmental way? How precise and convincing must results be so that authors call for significant policy changes? These questions are in general difficult to answer, because rhetoric on the one hand, and content and methodology of the paper on the other, cannot be separated easily. We, therefore, use a unique example to examine this question empirically: the analysis of gender wage differentials. Here, the Blinder-Oaxaca decomposition represents a standard research method that compares male and female earnings, holding productivity constant. We analyze close to 200 papers to investigate what drives authors to talk about "discrimination", whether and when they call for policy activism or when they are more hesitant to do so. Furthermore, we examine whether the rhetoric used really reveals an author's prejudice on the topic which may also be reflected in data selection and thereby his or her findings.

JEL Classification: J7, J16, B4

Keywords: rhetoric, gender wage differential, discrimination

Corresponding author:

Rudolf Winter-Ebmer
Department of Economics
University of Linz
4040 Linz
Austria
Tel.: +43 70 2468 8236
Fax: +43 70 2468 8238
Email: Rudolf.Winterebmer@jku.at

* Klaus Nowotny and Sandra Leitner provided invaluable assistance with the data collection. Thanks to seminar participants at the annual meeting of the European Society for Population Economics in New York as well as to René Böheim, Gabriel Felbermayr, Walter Ötsch and Andrea Weber for helpful comments. This research was supported by the Ludwig Boltzmann-Institut for Growth Research and the Austrian Science Funds (P15422).

1 Introduction

Economists like to think of themselves as purely rational, objective researchers who use scientific methods to gain quantifiable results. They claim to be unaffected by personal values in their research since scientific techniques would guarantee objective findings. However, what they say and do may be two different things. As McCloskey has put it "Economists do not follow the laws of enquiry their methodologies lay down" (1983; 168).

One the one hand, typically a whole number of scientifically acknowledged methods are available to the researcher and the choice of these might affect the result. One the other hand, there are different ways of presenting the same finding. In their dialogs, economists try to "persuade" each other of their hypotheses, models, or empirical results. For this they use their language, "the aptness of economic metaphors, the relevance of historical precedents, the persuasiveness of introspection, the power of authority, the charm of symmetry, the claims of morality" (McCloskey, 1983; 482).

For persuading others they often also make use of words and phrases which suggest certain connotations and interpretations; e.g. terms like efficiency, equilibrium, stability, maximization and the like which evoke scientific power (McCloskey, 1998). Even when discussing one's results there is a lot of room for interpretation. Particularly if there are no comparable studies an author might regard an estimated effect as rather big or "statistically significant but economically of second-order importance". (S)he can use suggestive or neutral terms when interpreting the result or can call for economic policies to fight apparent economic or social ills.

In contrast to previous papers on rhetoric in economics[1] we use a purely empirical research strategy. We examine a specific example in the economic literature: the calculation of the gender wage gap. The advantage of investigating a narrow topic such as the gender wage gap is that it restricts the analysis to papers which all tackle the same clear question: do equally productive females earn less than corresponding males? Furthermore, there exists a standardized research method, which has been internationally adopted to examine the gender wage differential. The Blinder (1973) and Oaxaca (1973) approach allows to decompose gender wage differentials into two parts: one part which is driven by different productivity (usually called "endowment effect" or "productivity effect") and the rest. Some researchers

[1] Discussions in the popular press on the rhetoric of economists is often concerned with its apparent dullness. The economic rhetoric is characterized as brimming with too many Greek-letters and figures (See Reeves, 2003 for a report on the Royal Economic Society Annual Meeting).

call the second part “discrimination effect”, others use the term "unexplained residual" because unobserved heterogeneity, i.e. the fact that men are more productive than women in their unobservable characteristics, might cause the results.[2] The term "sex discrimination" is more likely to be used by authors who believe that discrimination exists and is a social problem, while those who believe in the perfect functioning of free (input) markets will be more likely to speak of "unexplained residuals". Ideology, therefore, may affect the wording.

Generally, in the economic literature, the rhetoric of an author, the content and results of the analysis and the methods or methodology cannot be separated easily, because they are mutually dependent. Focusing on the gender wage gap, we avoid this problem since the Blinder-Oaxaca decomposition is such a standard procedure that it is used by anybody interested in the gender wage differential. Therefore, even though there are some variations to the Blinder-Oaxaca decomposition, the general method to analyze gender wage differentials is practically given.

Our empirical strategy is to look at published papers on the gender wage gap and explore their rhetoric, in particular the way how the authors describe their results. Section 2 quickly reviews the Blinder-Oaxaca decomposition technique which all the papers we investigate make use of, section 3 then describes our data set of previously published papers and the way we searched for our data. Section 4 analyzes the use of the word discrimination in the text and title of a research paper. Can we find any evidence, that attitudes and demographic characteristics of the author or the quality of research determine the terminology used? Rhetoric in describing one’s results is one thing, calling for (drastic) policy reforms is another – some might say “a more serious” thing. In the following section 5 we examine whether attitudes of authors and quality of the empirical estimate affect the occurrence and severity of policy recommendations to overcome the gender wage gap. Finally, in section 6 we look at more indirect forms of persuasion than rhetoric: in the course of an empirical project, there are many cross-roads one can take in terms of data selection, which might be reasonable and arguable, but which tend to lead the researcher to particular results. Section 7 concludes.

---

[2] It is interesting to note that both, Blinder (1973) as well as Oaxaca (1973), used the term discrimination.

2 Estimates for gender effects on wages

The standard procedure to investigate differences in wages is the one developed by Blinder (1973) and Oaxaca (1973). Wages are estimated separately for individuals i of the different groups g, males and females, which allows that productive characteristics of men and women are rewarded differently:

$$W_{gi} = b_g \; X_{gi} + e_{gi}, \tag{1.1}$$

g = (m, f) represents the two sexes; W_{gi} is the log wage and X_{gi} the control characteristics of an individual i of group g.

The total difference in mean wages of men and women can then be decomposed into two parts:

$$\overline{W_m} - \overline{W_f} = (\overline{X_m} - \overline{X_f})\hat{b}_m + (\hat{b}_m - \hat{b}_f)\overline{X_f} \;\equiv\; E + CE, \tag{1.2}$$

where $\overline{W_g}$ and $\overline{X_g}$ denote the mean log wages and productivity characteristics of group g and $\hat{b}_g$ represents the estimated parameters from equation (1.1). While the first term stands for the effect of different productive characteristics (the endowment effect E), the second term represents the gender effect which is due to differences in the estimated coefficients for both groups and is often referred to as "discrimination effect". Other authors emphasize that men may still be more productive than women even if unobservable in the data and refer to it as "unexplained residual". Others again, although the minority, try to avoid both terms by referring e.g. to a "price effect", "sex effect" or "unequal treatment". In the following we will refer to this estimated wage component CE as the "calculated effect" which authors present in their study.

Since the first use in the early seventies, hundreds of authors have adopted and also extended the Blinder-Oaxaca approach.[3] We investigated their papers to analyze the rhetoric they use when presenting their results. The word "unexplained residual" might be used either if an author does not believe in the existence of discrimination at all or if (s)he tries to leave the interpretation up to the reader. In principle, a positive unexplained residual can be reconciled with zero discrimination or with any amount of discrimination that lies between

[3] Later, some refinements of the decomposition technique have been introduced by e.g. Brown et al. (1980), Reimers (1983), Cotton (1988), and Neumark (1988). Papers using any of these techniques have also been considered in our analysis.

zero and the calculated effect.[4] In terms of persuasion, therefore, people who believe that sex discrimination exists and is a social ill will be more likely to name their empirical results "discrimination" to encourage the reader to interpret it accordingly. Others, who believe that markets correctly assess and reward market productivity will be more likely to use the term "unexplained residual" if they find differences in wages which cannot be attributed to observable personal characteristics.

3 Data

In November 2000, we searched the Economic Literature Index for any reference to: "(wage* or salar* or earning*) and (discrimination or differen*) and (sex or gender)". This search strategy led us to 1541 references. After examination of abstracts and articles we identified 192 empirical articles which used regression analysis and a Blinder-Oaxaca decomposition to analyze male-female wage differentials.[5]

Since we wanted to examine how the rhetoric of a paper was determined by variables like the calculated effect or the methods used, we took into account all the different published estimates of one paper which were based on different methods and data sets and used the mean values of these variables for analysis. Figure 1 shows the use of the term discrimination over time: in the early 1970s about two thirds of authors (papers) called the calculated effect discrimination, whereas only about 20% of authors did so in the late 1990s. A similar picture can be found in the titles of the papers. While more than 50% of papers used the word discrimination in the title in the beginning of our period, this rate dropped to less than 20% at the end (see Figure 2).

[4] In principle the discrimination effect could also be larger than the unexplained differential when men are doing worse in the unobservable than women, for example if they use drugs and are involved in illegitimate practices more frequently.

[5] A full list of papers included in this study can be downloaded from the following URL:
www.economics.uni-linz.ac.at/weichsel/work/rhetoric_papers.doc
Some empirical studies used regression analysis but no Blinder-Oaxaca decomposition. In total 70 articles included only sex dummies in the wage regressions and were excluded from our study. Since these papers typically did not focus on gender wage differentials, their use here would have contaminated our experimental setting.

4 Do authors use the term discrimination?

We investigated nearly 200 papers using the Blinder-Oaxaca decomposition and its derivatives to analyze the rhetoric authors use when presenting their results. As this decomposition technique and its refinements are well defined and standard in the literature, a researcher knows exactly what (s)he is talking about technically, although he or she may interpret the "calculated effect" CE differently. Some may consider a positive CE as a proof for discrimination, and conclude that the government should take measures to fight it. Others may be less convinced that they have found true "discrimination". In our analysis of whether wage differentials are assigned to discrimination we first only looked at the immediate discussion and interpretation of their calculated effect CE, not at the rhetoric the authors might have used in the introduction or conclusions of their paper. The reason is that also authors who do not believe in the existence of discrimination will still use the term throughout their paper to lay out their argument. A simple word count therefore would not be sufficient. One would have to evaluate whether the word discrimination was used in an affirmative or negative way and set the number of mentions in relation to the number of themes covered in the paper, which might be difficult to assess. The more clear-cut approach, however, is to simply analyze the discussion of the calculated effect since it is constitutional to the persuasive power of the argument. In a further analysis we also investigate the terminology in the title of a paper.[6]

Determinates of economic rhetoric

Which factors determine the terminology of an author? First candidates are beliefs, ideology and attitudes of the author. As direct information about the ideology of the researcher is not available in our case, we try to proxy the gender-related attitudes of the author by information about previous occupation with the topic. If an author has published on gender-related issues (in the last five years), we might assume, that he or she is more attached to the topic than others. The variable "number of previous gender papers" was created by searching for papers of the respective author via Econlit using the key-words "gender" or "sex" or "female" "women" (or woman) – going five years back. Additionally, we also include demographic information for being female or based at a U.S. institution.

[6] Chevalier and Hudson (2001) conduct a similar quantitative text analysis when counting the occurrences of intentional terms in one volume of the Journal of Finance.

A second candidate is the calculated effect itself. The bigger it is, the more an author might be convinced that (at least a part) must be due to discrimination and not due to characteristics unobservable in the data. Factors which capture the reliability of the calculated effect may also determine the interpretation of the results. The decomposition analysis is meant to calculate the wage gap between equally productive males and females, where males and females are artificially made "equal" by the econometrician. If the author remains less convinced about the validity of his or her comparison, (s)he may be hesitant to call the decomposition result discrimination. We constructed several variables for the reliability of the calculated effect.[7] The first quality measures are purely "study-based" and rely directly upon observable quality characteristics of the research paper. The variable "sound/elaborate method" measures whether a study used any of the more elaborate decomposition techniques as developed by Neumark, Reimers, Cotton, or Brown et al. "Good data", on the other hand, describes the quality of the data used. It captures whether the data provide hourly wages and actual work-experience, and also accounts for whether the data set is large (N > 1000) and comes from administrative sources.[8] In the latter case the earnings data should be more reliable. Additionally, if the authors presented more estimates of the gender wage gap and/or had a higher R^2 in the underlying wage regressions, we might consider the results as more trustworthy.

Furthermore, we included more general, "market-based" quality indicators in our study, capturing the quality of the research(er) via journal quality and previous publication record of the author. For the rank of a journal we used the citation-based journal rankings from Laband and Piette (1994).[9] The variable "number of previous papers in core journals" counts the number of articles an author had published in the previous five years in one of the core journals as defined by Stigler et al. (1995).[10]

---

[7] An obvious candidate would be to take the precision of the estimate (in general the standard error) as a quality indicator. However, this cannot be done in our case, because the precision of the calculated effect, the constructed indicator in the Blinder-Oaxaca decomposition, is usually not reported in the literature. See Silber and Weber (1999) for a bootstrap approach to construct standard errors for different decomposition procedures.

[8] The variable "good data" simply counts how many of the above mentioned characteristics apply.

[9] The 30 highest ranked journals were classified as "top journals", the following 40 journals as "medium journals".

[10] For multiple authors their mean values were taken. Stigler et al. (1995) define the following journals as "core journals": American Economic Review, Economic Journal, Econometrica, Journal of Economic Theory, Journal of Monetary Economics, Journal of Political Economy, Quarterly Journal of Economics, Review of Economics and Statis tics, Review of Economic Studies.

The use of terminology in the text

The results of our probit-analysis for the use of the term “discrimination” in the course of the discussion of the calculated effect are reported in Table (1).[11] The first column includes only our proxy for attitudes and demographic characteristics of the authors, together with a time trend. It turns out that women are equally inclined to use the term “discrimination” as are men. However if they are co-authoring the paper with a male author; the probability to refer to discrimination is 20 percentage points lower. It might be the case, that such a mixed-sex pair is more conservative in its wording than other authors; they also may find it more difficult to agree on a joint rhetoric for their paper. However, this negative co-author effect does not apply to same-sex pairs. Authors who have a pronounced interest in the gender topic – shown by previous publications– do not use a different rhetoric. Authors from U.S. universities again use the term discrimination 24 percentage points less frequently. Also, over the years it seems to have become less popular to speak of discrimination.

The next columns investigate the effect of quality characteristics of the research paper which reflect the reliability of the calculated gender wage gap. The coefficient for the calculated effect itself is positive as expected: the bigger the calculated difference the more likely authors are to use the word "discrimination". If the calculated effect is big, authors are more convinced that at least part of their result must be due to discrimination and not due to the neglect of an unobservable variable.[12] We also find that top journals avoid the term "discrimination" more often.[13] Including the research history of authors in our set of explanatory variables, we find that the more papers an author has previously published in core journals the less likely (s)he is to use the word “discrimination”. The effect is rather strong, one article reducing the probability by 20-30 percentage points. While “quality” of the author and the journal seem to indicate less willingness to refer to "discrimination", more direct, "study-based" indicators about the reliability of the research point in the opposite direction. The use of more sophisticated decomposition techniques, the availability of better data and – most prominently – a good explanation of wages via a high R^2 in the original wage regressions, all these factors increase the use of the term “discrimination” considerably and

[11] In Table (5) in the Appendix we extend our analysis and distinguish between three categories: papers which call their calculated effects "unexplained residuals", "discrimination-effect" or use some other term, e.g. price or sex effect, unequal treatment, or salary differential. "Discrimination" is hereby considered the strongest expression - used in 41 % of papers, "unexplained residual" the weakest (used in 42 % of papers). The signs of the coefficients remain the same, but the level of significance somewhat decreases.

[12] Recall that the calculated effect (CE) represents the log wage differential as defined in (1.2). In our sample it has a mean of 0.193 and a standard deviation of 0.13.

with sizable quantitative effects. Column (4) includes attitude and demographic characteristics as well as quality indicators to find the impact of the attitude and demographic indicators shrink somewhat. Column (5) indicates that a high calculated effect only leads the author to refer to "discrimination", if (s)he has also good data at disposal, as is shown by the interaction effect. These results indicate that the quality of the researcher and the research outlet has a different effect than the study-specific quality (reliability) of the actual paper. While the standard of the researcher and the outlet call for a more cautious and conservative wording, reliability of the research paper at hand calls for a more determined and pronounced standpoint.

The use of terminology in the title

Apart from the rhetoric used in the text, we are also interested whether authors use the term discrimination in the title of their papers. Note, that the current experiment is less clear than the use of discrimination in the interpretation of the Blinder-Oaxaca decomposition. The title of the paper might relate to the potentially large number of themes covered in a paper; it might have questions in it (e.g. Is there sex discrimination in Chile?); or the author might simply want to relate the title to a thread of research papers. All this makes the interpretation of the title more difficult.

The results of our probit analysis are presented in Table (2). As expected, the results are less precise than in Table (1). Two variables have a strong and consistent association with the probability of using the term "discrimination": sex and country of residence. We consistently find that women and U.S. authors more often abstain from the use of the word discrimination in the title of their papers. On the other hand, authors who used more sophisticated decomposition techniques, speak of "discrimination" more often. Since we used the full specifications from Table (1), we can also see which variables had *no* impact on the rhetoric in the title: neither the size of the calculated coefficient, previous occupation with the topic nor the quality of the journal had any impact on the wording.

[13] Cherry and Feiner (1992) find that those journals abstracted in the Journal of Economic Literature, i.e. the more prestigious journals, reduced the number of articles on discrimination between 1972 and 1987, while the total number of papers on the topic remained roughly the same.

5 When do authors give policy recommendations?

Since there seem to be major differences in the rhetoric of authors, which partly seem to rest upon author-specific characteristics, the next step will be to see whether such differences also appear with respect to policy recommendations. For example one could assume that people of a certain ideology, e.g. more liberal individuals, would be more likely than conservatives to assign differences in wages to discrimination. These authors then might also be more likely to advocate policy measures to combat labor market discrimination. Our prime interest is to understand, why authors call for policy intervention. Is it because they are particularly convinced of their empirical result (e.g. due to an exceptionally big calculated effect) or is it because of their biased attitudes and values?[14]

We constructed three different indicators for policy recommendations. The variable *"severity"* captures how strong the proposed state intervention would be. It ranges from the explicit rejection to adopt measures (-1), over no recommendation (0) and rather general advise (e.g. "remove barriers for women", "encourage college choice of females", "discourage sex-discrimination", (1)), to explicit laws that either prohibit discrimination (e.g. equal pay act, anti-discrimination law, (2)) or actively promote female employment (e.g. affirmative action, (3)). The variable *"determination of policy recommendation"* is an index for how determined and convinced an author is that a proposed measure should be realized and would be successful. It captures whether an author argues a measure "may, might, could" (1)/ "should, can" (2)/ "will, would" (3) lead to particular consequences, or how necessary (s)he states it is.[15] *"Number of words"* simply counts how many words a paper spends on policy recommendation. In total, 72 papers made any kind of policy proposal.

In our regression models we included all previously mentioned regressors and added a dummy variable, whether in the paper the calculated effect was attributed to discrimination. The reason for including this variable was firstly that assigning differences in wages to discrimination might be a precondition to ask for policy interventions (if authors believe that men and women are different in unobservables by nature, this might not be open to change). Secondly, this dummy variable acts as a proxy for the ideological predisposition of an author.

---

[14] See Fuchs, Krueger and Poterba (1998) on survey results on the importance of values for policy descriptions of labor and public economists.

[15] No suggestion of a measure is coded as 0. If an author included any additional reservation (e.g. the measure will *possibly* lead to…), this was accounted for by assigning a one point lower value.

Our different measures for policy conclusions lead to rather similar results. The main message of Table (3) is our inability to explain the variation in policy advice among authors by demographic characteristics, our indicators for their values and attitudes or by the quality and reliability of their research: neither female authors, nor those with a longer history in gender-related research, nor those having used the term discrimination in the interpretation of the Blinder-Oaxaca decomposition require more policy changes. This finding stands in contrast to Fuchs et al. (1998) who find that values of economists strongly affect their general policy positions. However, they also do not find a significant impact of (guessed) coefficient estimates on policy recommendations which corresponds to our results. This finding can be seen as reassuring if one accepts our indicators for underlying attitudes of the authors: Gender, ideology and values do not seem to influence policy advice of economists. Even among those authors who have been shown to interpret the calculated gender wage gap in a certain way – by calling it discrimination or unexplained residual – there is no difference in policy recommendations. This means that there are differences in the rhetoric of economists, but these do not explain the policy prescriptions of authors. However, one could argue that our indicators do not adequately capture ideology, and therefore the variation in the prevalence of giving policy advice cannot fully be explained.

We also find no evidence of the impact of quality of the research and research outlet on policy recommendations. Only authors who used more sophisticated methods are more convinced about their own research output and use more words in their policy conclusions; those who presented a larger number of empirical results in their paper are, in turn, more reserved in giving advice. The number of authors a paper has, significantly reduces the number of words spent on policy recommendations. Probably, again, it is harder for a larger group of writers to reach a consensus. On the other hand, the number of words spent on policy as well as the determination of authors in their policy conclusions seem to increase over time. This is interesting if one considers previous results where we showed that the use of the term discrimination decreased over the years.

6 Restrictions to data set

The gender wage gap differs greatly for different subgroups within a population. For example for new entries in the labor market we typically observe lower differences in wages. The same is true for the public sector and for never-married individuals. If wage differentials are

calculated for individuals within a rather narrowly defined occupation the gender wage gap is typically lower than when people of all different job types are investigated. Also women in high prestige occupations are typically confronted with a smaller pay gap than those in low prestige occupation.[16]

Some data sets only cover a very specific group of individuals, e.g. exist for a specific occupation only or for a particular group of people (like new entries). Consequently, the choice of data a researcher uses will crucially affect the calculated gender wage gap. Obviously, the reader of a paper which is based on a restricted data set will usually be aware that the applicability of the calculated outcome is restricted to the investigated subgroup and should interpret the results accordingly. However, in terms of "persuasion", choosing a restricted data set, which leads to a lower calculated effect than would be obtained for an entire sample of the population, tends to create the impression that differences in earnings between the sexes are a minor economic problem.[17] Therefore, researchers who want to convince their audience either about the importance or irrelevance of discrimination may base their analysis on selective data sets.

For our analysis we created a variable that counts the number of data set restrictions that a research paper is based on. The number of restrictions which increase the gender wage gap entered this variable additively; restrictions decreasing the gap were subtracted.[18] If authors used representative data but calculated effects for different subgroups of an entire population in their paper this was not considered a restriction, since a full picture of the entire population was presented. 25 % of all investigated papers were based on at least one - up to three - restrictions generally thought as reducing the gender wage gap. Only 4 % of authors used one restriction which increases the gap compared to the entire population.[19]

The goal of our investigation was to examine whether attitudes or research quality were correlated with an author's choice of data. In Table 4 we examine the use of restricted data sets via an ordered probit model. In column (1) we only include proxies for attitudes

---

[16] See Weichselbaumer and Winter-Ebmer (2003) for a meta-analysis of gender wage gap studies, who show these regularities.

[17] See DeMarzo et al. (2003) for a formal bounded rationality model of opinion formation, where people have a persuasion bias. They argue that readers should in principle make the right discount for – in our case – a selected data set, but often they fail to do so; especially if they get a particular message repeatedly. Without such a persuasion bias, many phenomena in political and economic marketing could not be explained.

[18] As restrictions lowering the gender wage gap we considered the following subgroups: new-entries, never married individuals, workers in the public sector, in a specific narrow occupation and in high-prestige occupations. Only investigating the private sector, low prestige occupations and married people was considered as increasing the wage differential.

[19] It should be noted, that using a restricted data set may be good concerning the reliability of the research result, because individuals are better comparable within a subset of the population. However, here we investigate

(including rhetoric) as well as demographic variables of the author: neither gender nor previous occupation with gender topics are related to the use of selective data sets. Moreover, our indicators for attitudes, like the use of discrimination to describe the gender wage gap as well as the quest for policy conclusions[20], are unrelated to the choice of data restrictions. These results are reassuring again: Authors might differ in their assessment and their rhetoric describing a phenomenon, but this does not influence important decisions in the course of the empirical work. Over our observed time span from the 70s to the 90s, however, authors reduced their use of restrictions which tend to lower the gender wage gap.

In column (2), only indicators for quality of research paper and author are included with some control variables. We find that papers which restrict their analysis to a subset of a population with a lower gender wage gap are more often published in top and medium ranked journals. This correlation may be due to editorial policy or simply to the quality of the respective papers. A narrower data set might be seen as a quality indicator as it makes the essential problem of the calculation of the gender wage gap easier and compares likes with likes. Column (3) provides a full specification which confirms previous results.

7 Conclusions

Economists, when publishing academic papers, compete with another for the attention of readers and policy-makers. Naturally, the authors' rhetoric is a valuable tool to persuade the audience of the importance and accuracy of one's own approach and results. In this paper, we concentrate on the interaction between attitudes, assessment of research results and rhetoric. Economists can present their research and their results in many different ways, neutral or judgmental; they might draw policy conclusions from their results or not. All of this might have consequences on the public reception of the paper in the science community, but also in policy circles.

In contrast to hermeneutic text analysis as practiced in literature[21] and in the science of history affected by the so-called cultural turn[22], this paper's consideration of rhetoric has been solely quantitative. However, this may be the language economists understand best, anyway.

---

whether we find an indication that data sets might have been chosen for reasons of persuasion to confirm authors' values and expectations.

[20] When substituting for other indicators for policy conclusions these were insignificant likewise.

[21] For an overview on how economic texts are examined for their form, content and contexts see Woodmansee and Osteen (1999).

We use the example of the rhetoric of the gender wage gap, because the method of the Blinder-Oaxaca decomposition is universally used in labor economics: we identified almost 200 papers that have applied it, which makes a quantitative analysis of rhetoric possible. In particular, we concentrated on the use of the word "discrimination" and the tendency to draw policy conclusions from the analysis.

What are the results?

Interestingly, females and U.S. authors both use a less explicit terminology and are more reluctant to use the word "discrimination" when investigating the gender wage gap; they may be more cautious in interpreting a given phenomenon in a contentious way. If the calculated gender wage gap is lower or the results appear less reliable, authors refrain from strong wording. Likewise, authors in top journals use it less often to show prudent and scientifically sound behavior.

On the other hand, policy conclusions economists make are not affected by attitudes and values of authors. Irrespective of the author's views about discrimination, the prevalence and severity of policy conclusions are the same. Likewise, attitudes of authors about discrimination do not influence their research strategies. There is no indication that authors who believe that no discrimination exists choose a particular data set which would be likely to lead to a lower gender wage gap. These results are

We find that females and U.S. economists use the term "discrimination" less often when referring to the gender wage gap. Since U.S. economists may have stronger believes in the efficiency of markets than e.g. Europeans, it does not come as a surprise that U.S. authors use a more conservative terminology. Women, on the other hand, may be less pronounced in their wording to avoid being considered preoccupied and in order to get their research published in male-dominated journals. While authors finding high gender wage gaps or having used more reliable data or methods speak of "discrimination" more often, the contrary is true when an article is published in a top journal. This may be caused by a more conservative editorial policy of these journals.

Concerning policy recommend-dations, demographic variables and objective information about the content, methodology and results of the research cannot explain the variance in the

[22] For new history of science see Dennis (1997).

very comforting for the economics profession in the sense that while the rhetoric of individuals may differ somewhat, decisions concerning research design and policy advice are not influenced at all.

prevalence and severity of policy conclusions. There must be some – unobservable but fixed – factors of ideology which are shaping policy advice irrespective of the actual result the author has found. Moreover, the same unobservable values seem to drive empirical methodology in the way how authors construct their data sets in order to fit their predetermined views.

8 References

Blinder, Alan S.: Wage Discrimination: Reduced Form and Structural Estimates, Journal of Human Resources, 8/4, 1973, 436-455.

Brown, Randall S., Moon, Marylin and Barbara S. Zoloth: Incorporating Occupational Attainment in Studies of Male-Female Earnings Differentials, Journal of Human Resources, 40/1, 1980, 3-28.

Cherry, Robert and Susan Feiner: The Treatment of Racial and Sexual Discrimination in Economics Journals and Economics Textbooks: 1972 to 1987, Review of Black Political Economy, 1992, 99-118.

Chevalier, Gisèle and Richard Hudson: The Use of Intentional Language in Scientific Articles in Finance, Journal of Economic Methodology, 8/2, 2001, 203-228.

Cotton, Jeremiah: On the Decomposition of Wage Differentials, Review of Economics and Statistics, 70/2, 1988, 236-243.

DeMarzo, Peter, M., Dimitri Vayanos and Jeffrey Zwiebel: Persuasion Bias, Social Influence, and Unidimensional Opinions, Quarterly Journal of Economics 115/3, 2003, 910-968.

Dennis, Michael Aaron: Historiography of Science. An American Perspective, in: Krige, John and Dominique Pestre (eds.), Science in the Twentieth Century, Harwood Academic Publishers, Amsterdam, 1997, 1-26.

Fuchs, Victor R., Alan B. Krueger and James M. Poterba: Economists' Views about Parameters, Values, and Policies: Survey Results in Labor and Public Economics, Journal of Economic Literature 36, 1998, 1387-1425.

Laband, David N. and Michael J. Piette: The Relative Impacts of Economics Journals: 1970-1990, Journal of Economic Literature, 32/2, 1994, 640-667.

McCloskey, Donald: The Rhetoric of Economics, Journal of Economic Literature, 21, 1983, 481-517.

McCloskey, Deirdre: The Rhetoric of Economics, second edition, University of Wisconsin Press, Madison, 1998.

Neumark, David: Employers' Discriminatory Behavior and the Estimation of Wage Discrimination, The Journal of Human Resources, 23/3, 1988, 279-95.

Oaxaca, Ronald: Male-Female Wage Differentials in Urban Labor Markets, International Economic Review, 14/3, 1973, 693-709.

Reeves, Richard: Conference Report on the RES conference 2003, Newsletter of the Royal Economic Society, 112, July 2003, 5-8.

Reimers, Cordelia W.: Labor Market Discrimination against Hispanic and Black Men, Review of Economics and Statistics, 65/4, 1983, 570-579.

Silber, Jacques and Michal Weber: Labor Market Discrimination: Are there Significant Differences between the Various Decomposition Procedures?, Applied Economics 31, 1999, 359-365.

Stanley, T. D. and Stephen B. Jarrell: Gender Wage Discrimination Bias? A Meta-Regression Analysis, Journal of Human Resources, 33/4, 1998, 947-973.

Stigler, George J., Stephen M. Stigler, and Clair Friedland: The Journals of Economics, Journal of Political Economy, 103/2, 1995, 331-359.

Weichselbaumer, Doris and Rudolf Winter-Ebmer: A Meta-Analysis on the International Gender Wage Gap, mimeo, University of Linz, Austria, 2003.

Woodmansee, Martha and Mark Osteen (eds): The New Economic Criticism. Studies at the Intersection of Literature and Economics, Routledge, New York, 1999.

9 Tables

Table 1: Probability to refer to calculated effect as "discrimination" in text (marginal effects)

	(1)	(2)	(3)	(4)	(5)
Female coauthor	-0.214			-0.160	-0.170
	(0.083)			(0.103)	(0.103)
Female authorship	-0.010			0.051	0.033
	(0.094)			(0.115)	(0.117)
US author(s)	-0.244			-0.158	-0.163
	(0.077)			(0.093)	(0.095)
# of previous gender papers	-0.028			0.006	0.002
	(0.018)			(0.022)	(0.022)
# of authors	-0.038	-0.074	-0.073	-0.009	-0.025
	(0.064)	(0.057)	(0.068)	(0.079)	(0.081)
Year	-0.013	-0.019	-0.021	-0.020	-0.020
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Calculated effect		0.735	0.803	0.770	-0.060
		(0.315)	(0.385)	(0.396)	(0.662)
Top journal		-0.282	-0.346	-0.327	-0.326
		(0.068)	(0.068)	(0.071)	(0.073)
Medium journal		-0.092	-0.164	-0.157	-0.164
		(0.093)	(0.098)	(0.099)	(0.099)
# of prev. papers in core journals		-0.324	-0.386	-0.386	-0.377
		(0.101)	(0.126)	(0.140)	(0.138)
Sophisticated method (0 - 1)		0.209	0.199	0.201	0.196
		(0.091)	(0.106)	(0.109)	(0.110)
Good data (0 - 4)		0.161	0.184	0.170	-0.034
		(0.053)	(0.061)	(0.063)	(0.137)
# of estimates per study		0.029	0.030	0.035	0.031
		(0.021)	(0.024)	(0.025)	(0.025)
R^2 of wage reg.			0.670	0.687	0.583
			(0.285)	(0.296)	(0.307)
Calculated Effect * good data					1.001
					(0.619)
Observations	191	192	160	159	159
Pseudo R^2	0.10	0.19	0.22	0.25	0.26

Standard errors in parentheses

Table 2: Probability to use term "discrimination" in title of paper (marginal effects)

	(1)	(2)	(3)	(4)	(5)
Female coauthor	-0.051			-0.048	-0.050
	(0.078)			(0.094)	(0.093)
Female authorship	-0.170			-0.149	-0.154
	(0.059)			(0.076)	(0.075)
US author(s)	-0.127			-0.146	-0.148
	(0.065)			(0.080)	(0.080)
# of previous gender papers	-0.006			0.012	0.011
	(0.014)			(0.019)	(0.019)
# of authors	-0.100	-0.062	-0.067	-0.089	-0.095
	(0.058)	(0.049)	(0.059)	(0.068)	(0.069)
Year	-0.009	-0.012	-0.009	-0.007	-0.008
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
Calculated effect		0.059	0.033	-0.007	-0.331
		(0.268)	(0.323)	(0.334)	(0.569)
Top journal		-0.025	-0.005	0.021	0.021
		(0.089)	(0.111)	(0.120)	(0.120)
Medium journal		-0.019	0.011	0.027	0.024
		(0.088)	(0.106)	(0.111)	(0.110)
# of prev. papers in core journals		-0.112	-0.098	-0.113	-0.110
		(0.070)	(0.078)	(0.088)	(0.086)
Sophisticated Method (0 - 1)		0.162	0.199	0.167	0.165
		(0.078)	(0.092)	(0.092)	(0.092)
Good data (0 - 4)		0.015	0.019	-0.006	-0.078
		(0.043)	(0.049)	(0.050)	(0.112)
# of estimates per study		0.019	0.025	0.024	0.022
		(0.016)	(0.019)	(0.020)	(0.020)
R^2 in wage reg.			0.334	0.384	0.344
			(0.232)	(0.241)	(0.246)
Calculated effect * good data					0.353
					(0.494)
Observations	191	192	160	159	159
Pseudo R^2	0.08	0.08	0.08	0.12	0.12

Standard errors in parentheses

Table 3: Policy recommendations

	Severity of policy recommendation		Determination of policy recommendation		# of words of policy recommendation	
	Ordered Probit		Ordered Probit		Tobit	
"Discrimination" in text	-0.150	-0.227	-0.113	-0.018	-6.128	-4.392
	(0.226)	(0.204)	(0.240)	(0.219)	(13.875)	(12.763)
Calculated effect	0.772	0.299	1.002	0.951	25.924	23.416
	(0.896)	(0.792)	(0.976)	(0.877)	(56.262)	(51.633)
Year	0.018	0.002	0.039	0.044	2.788	2.972
	(0.019)	(0.017)	(0.021)	(0.019)	(1.220)	(1.148)
Female coauthor	0.176	0.312	0.199	-0.039	15.600	9.237
	(0.285)	(0.255)	(0.308)	(0.277)	(17.833)	(16.020)
Female authorship	0.198	0.285	0.315	0.221	7.282	2.937
	(0.265)	(0.246)	(0.275)	(0.258)	(16.000)	(15.248)
US author(s)	0.060	-0.003	0.146	0.028	19.658	15.231
	(0.226)	(0.199)	(0.238)	(0.212)	(13.712)	(12.318)
# of authors	-0.401	-0.403	-0.549	-0.420	-36.607	-33.261
	(0.194)	(0.172)	(0.219)	(0.194)	(13.300)	(11.866)
Top journal	-0.175	-0.238	-0.329	0.008	-31.883	-16.429
	(0.353)	(0.291)	(0.403)	(0.324)	(23.911)	(19.513)
Medium journal	0.070	0.019	-0.113	0.016	-16.135	0.338
	(0.302)	(0.266)	(0.328)	(0.291)	(19.481)	(16.839)
# of prev. papers in core journals	0.077	0.083	-0.013	0.019	-3.631	-3.262
	(0.164)	(0.146)	(0.177)	(0.158)	(10.683)	(9.399)
# of previous gender papers	-0.060	-0.087	-0.060	-0.043	-3.680	-0.003
	(0.058)	(0.048)	(0.072)	(0.055)	(4.239)	(2.957)
Sophisticated Method (0 - 1)	0.316	0.341	0.371	0.453	27.952	34.372
	(0.243)	(0.216)	(0.260)	(0.232)	(15.003)	(13.576)
Good data (0 - 4)	0.034	-0.039	0.045	-0.022	9.079	6.086
	(0.137)	(0.124)	(0.145)	(0.133)	(8.468)	(7.842)
# of estimates per study	-0.097	-0.073	-0.155	-0.162	-6.354	-7.613
	(0.058)	(0.051)	(0.064)	(0.059)	(3.610)	(3.380)
R^2 of wage reg.	-0.071		-0.130		-54.313	
	(0.690)		(0.745)		(43.269)	
Observations	160	192	160	192	160	192
Pseudo R^2	0.05	0.05	0.08	0.07	0.04	0.03

Standard errors in parentheses

Table 4: Use of data restrictions (Ordered Probit)

	(1)	(2)	(3)
"Discrimination"	0.339		0.292
in text	(0.194)		(0.199)
# of words of policy	0.001		0.001
	(0.003)		(0.003)
Female coauthor	0.096		0.049
	(0.238)		(0.242)
Female authorship	0.210		0.227
	(0.234)		(0.242)
US author(s)	-0.184		-0.131
	(0.188)		(0.200)
# of previous gender	0.058		0.069
Papers	(0.040)		(0.046)
# of authors	-0.096	-0.072	-0.033
	(0.157)	(0.131)	(0.162)
Year	0.037	0.029	0.033
	(0.017)	(0.016)	(0.017)
Top journal		-0.615	-0.548
		(0.258)	(0.266)
Medium journal		-0.546	-0.576
		(0.247)	(0.252)
# of previous papers		0.060	0.031
in core journals		(0.117)	(0.143)
Observations	192	192	192
Pseudo R^2	0.04	0.04	0.06

Standard errors in parentheses

10 Figures

Figure 1

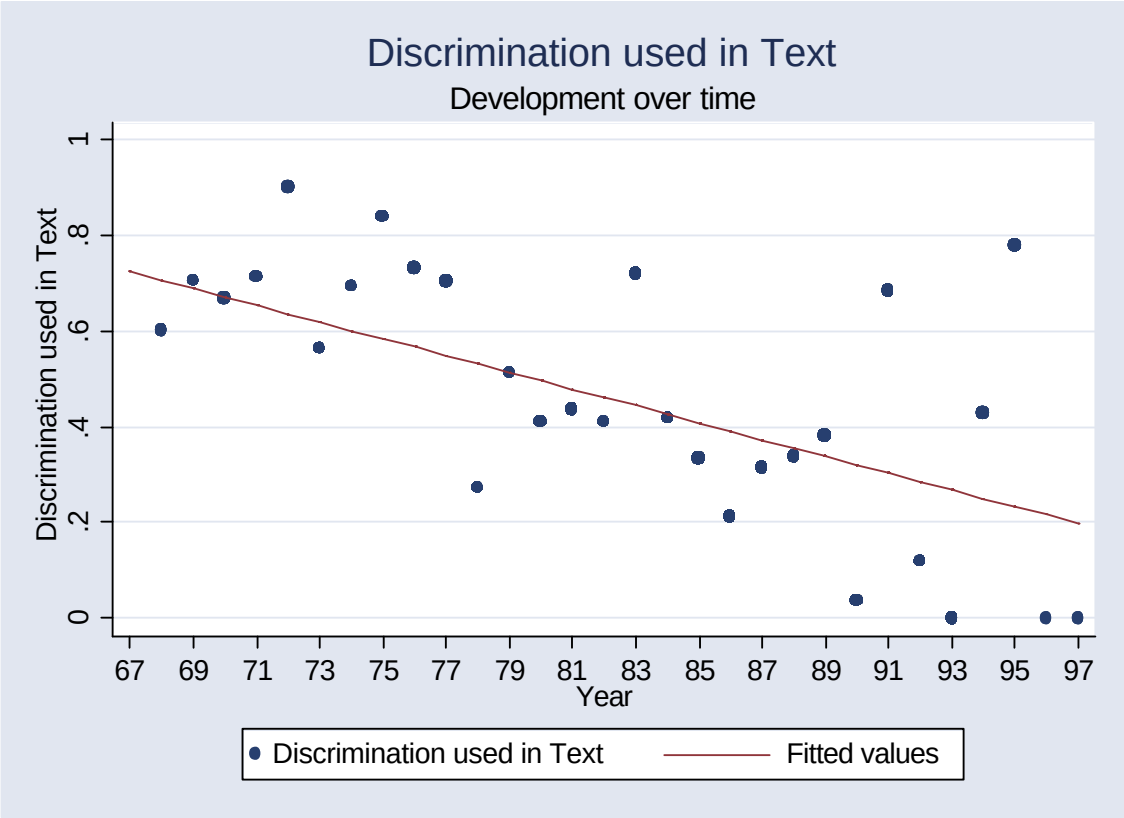
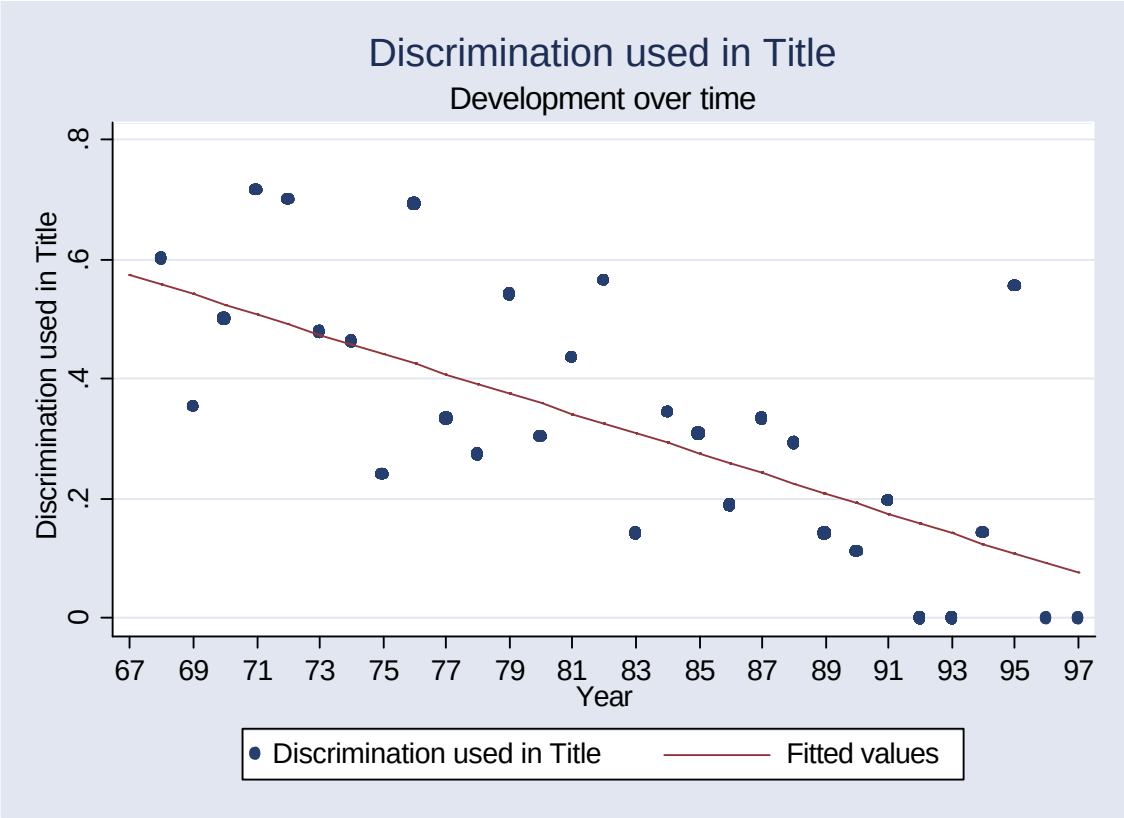


Figure 2


Discrimination used in Title
Development over time
Discrimination used in Title
0
.2
.4
.6
.8
67
69
71
73
75
77
79
81
83
85
87
89
91
93
95
97
Year
Discrimination used in Title
Fitted values

11 Appendix

Table 5: Calculated effect referred to as discrimination, sex-effect etc., or unexplained residual (Ordered Probit)

	(1)	(2)	(3)	(4)	(5)
Female coauthor	-0.472			-0.199	-0.240
	(0.232)			(0.268)	(0.270)
Female authorship	0.226			0.360	0.326
	(0.225)			(0.262)	(0.266)
US author(s)	-0.487			-0.426	-0.425
	(0.181)			(0.216)	(0.217)
# of previous gender papers	-0.070			-0.011	-0.019
	(0.041)			(0.053)	(0.053)
Year	-0.026	-0.030	-0.032	-0.033	-0.035
	(0.016)	(0.016)	(0.018)	(0.019)	(0.019)
# of authors	0.006	-0.143	-0.075	0.052	0.023
	(0.152)	(0.132)	(0.150)	(0.180)	(0.180)
Calculated effect		0.910	0.958	0.806	-1.764
		(0.750)	(0.855)	(0.877)	(1.415)
Top journal		-0.814	-1.101	-0.960	-0.968
		(0.286)	(0.338)	(0.349)	(0.353)
Medium journal		-0.015	-0.180	-0.194	-0.223
		(0.246)	(0.272)	(0.278)	(0.280)
# of prev. papers in core journals		-0.335	-0.338	-0.244	-0.281
		(0.130)	(0.148)	(0.161)	(0.163)
Sophisticated method		0.145	0.072	0.104	0.075
		(0.206)	(0.230)	(0.237)	(0.238)
Good data		0.273	0.308	0.266	-0.372
		(0.120)	(0.132)	(0.134)	(0.296)
# of estimates per study		0.080	0.096	0.116	0.099
		(0.049)	(0.055)	(0.057)	(0.058)
R^2 of wage reg.			1.592	1.614	1.205
			(0.650)	(0.675)	(0.702)
Calculated effect * good data					3.246
					(1.369)
Observations	192	192	160	160	160
Pseudo R^2	0.05	0.07	0.09	0.11	0.12

Standard errors in parentheses

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
891	L. Flood E. Pylkkänen R. Wahlberg	From Welfare to Work: Evaluating a Proposed Tax and Benefit Reform Targeted at Single Mothers in Sweden	6	10/03
892	B. T. Hirsch	What Do Unions Do for Economic Performance?	5	10/03
893	K. Sabirianova Peter	Skill-Biased Transition: The Role of Markets, Institutions, and Technological Change	4	10/03
894	R. Winkelmann	Parental Separation and Well-Being of Youths	7	10/03
895	J. M. Fitzgerald D. C. Ribar	Transitions in Welfare Participation and Female Headship	3	10/03
896	S. W. Polachek	What Can We Learn About the Decline in U.S. Union Membership from International Data?	2	10/03
897	M. Brown A. Falk E. Fehr	Relational Contracts and the Nature of Market Interactions	7	10/03
898	G. J. van den Berg A. G. C. van Lomwel J. C. van Ours	Nonparametric Estimation of a Dependent Competing Risks Model for Unemployment Durations	1	10/03
899	M. Karanassou H. Sala D. J. Snower	Unemployment in the European Union: Institutions, Prices, and Growth	2	10/03
900	M. C. Berger G. C. Blomquist K. Sabirianova Peter	Compensating Differentials in Emerging Labor and Housing Markets: Estimates of Quality of Life in Russian Cities	4	10/03
901	P. F. Orazem M. Vodopivec	Do Market Pressures Induce Economic Efficiency: The Case of Slovenian Manufacturing, 1994-2001	4	10/03
902	P. Frijters M. A. Shields S. Wheatley Price	Immigrant Job Search in the UK: Evidence from Panel Data	1	10/03
903	S. Girma H. Görg	Evaluating the Causal Effects of Foreign Acquisition on Domestic Skilled and Unskilled Wages	2	10/03
904	S. P. Jenkins P. Van Kerm	Trends in Income Inequality, Pro-Poor Income Growth and Income Mobility	1	10/03
905	D. Weichselbaumer R. Winter-Ebmer	Rhetoric in Economic Research: The Case of Gender Wage Differentials	7	10/03

An updated list of IZA Discussion Papers is available on the center‘s homepage www.iza.org.