A META-ANALYSIS OF THE INTERNATIONAL GENDER WAGE GAP

Doris Weichselbaumer

University of Linz and Universitat Pompeu Fabra

Rudolf Winter-Ebmer

University of Linz and Institute for Advanced Studies, Vienna

Abstract. Since the early 1970s, a number of authors have calculated gender wage differentials between women and men of equal productivity. This meta-study provides a new quantitative review of this vast amount of empirical literature on gender wage differentials as it concerns not only differences in methodology, data, and time periods, but also different countries. We place particular emphasis on a proper consideration of the quality of the underlying study which is done by a weighting with quality indicators. The results show that data restrictions – i.e. the limitation of the analysis to new entrants, never-marrieds, or one narrow occupation only – have the biggest impact on the resulting gender wage gap. Moreover, we are able to show what effect a misspecification of the underlying wage equation – like the frequent use of potential experience – has on the calculated gender wage gap. Over time, raw wage differentials worldwide have fallen substantially; however, most of this decrease is due to better labor market endowments of females.

Keywords. Gender wage differential; Meta-analysis

1. Introduction

The literature on the economics of discrimination started with Becker's seminal study in 1957. Since then – due to the proliferation of the use of microdata in the last three decades – the study of gender wage differentials became a routine job for labor economists. Microdata allowed to assess the productivity of individuals and to compare wages of equally productive males and females. In particular, the decomposition technique – as pioneered by Blinder (1973) and Oaxaca (1973) – has frequently been applied to data from the most different countries and time periods.

Given the importance and timeliness of the topic, many reviews or surveys of the development of gender wage gaps have been done.[1] Most of them concentrated on single countries and on econometric issues and were of a narrative type. With regard to the sheer number of available studies, any narrative survey will

have difficulties to condense and interpret these papers satisfactorily. Stanley and Jarrell (1998) as well as Jarrell and Stanley (2004) were the first to complement this survey literature with meta-analyses which systematically covered the published papers on gender wage differentials in the US. In their first study, in 1998, they identified 12 factors which affected the reported US gender wage differential and explained 80% of its variation. In their second study, in 2004, they revise their original paper including all additional US papers which almost doubles their entire data set.

In this paper, we extend their work and conduct a meta-analysis on the gender wage gap on the worldwide level. Furthermore, we place a particular emphasis on the consideration of the quality of the underlying study, which is done by a weighting with quality indicators, and examine the effects of data set restrictions.

Section 2 of the paper briefly discusses the method of meta-analysis and draws attention to some advantages and caveats with respect to this method. Section 3 shortly reviews the way gender wage differentials are calculated, while Section 4 discusses our data-generation process – a very important step in any meta-analysis. Section 5 introduces our meta-regression model and discusses problems some of which will be addressed by a weighting mechanism. Section 6 presents the results and Section 7 concludes.

2. Meta-Analysis

Meta-analysis is a helpful tool to cumulate, review, and evaluate empirical research. Papers investigating one particular topic are collected and analyzed concerning their data and method. Meta-analysis then allows evaluating the effect of different data characteristics and methodologies on the result reported, e.g. a regression parameter (Stanley, 2001). Instead of the usual practice of analyzing observations of individual workers, here, each previously conducted study represents one data point. Meta-regression analysis, in turn, uses regression techniques to explain these collected parameters by characteristics of the individual study.

One of the prime advantages of a meta-study over a narrative or a vote-counting review is that it allows a quantitative assessment of the literature in 'a way an econometrician would write a survey'. It offers a clear and systematic way to assess the merits of different research methods: all methodological features of a particular original study can be used as control variables in the meta-regression analysis; the resulting regression coefficients then give a quantitative measure of the importance of the concerned research methods. As meta-analysis is 'constructing' its own meta-data, the principle of completeness and replicability must dictate the choice of original papers. This implies that all papers have to be treated in a standardized way and there is no room for the reviewer for an individual assessment of papers. Typically in a narrative or vote-counting review, some papers are discarded due to methodological shortcomings, unreliability of the data and the like; on the other hand, some papers are highlighted. Obviously, the inclusion or exclusion of a paper lies in the personal assessment of the author. This can lead to discussions about the legitimacy of the choice of papers.[2] Meta-analysis avoids this problem as it includes all papers. However, differences

in the reliability of these original studies should not be disregarded. Therefore, in our meta-analysis, we developed some objective and operational indicators for the quality of a paper on the gender wage gap which are used as different weights in our meta-regression.

3. Estimates for Wage Differentials

The most common way to analyze discrimination based on gender is to compare male and female earnings holding productivity constant. One method is to simply include a sex dummy in the wage regression model:

$$W_i = \beta X_i + \gamma \text{sex}_i + \varepsilon_i, \tag{1}$$

where W_i represents the log wage and X_i the control characteristics (e.g. education, job experience, marital status, and job characteristics) of an individual. i, β, and γ are parameters.

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

$$W_{gi} = \beta_g X_{gi} + \varepsilon_{gi}, \tag{2}$$

where $g = (m, f)$, 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 wage differential between men and women can then be decomposed into an explained part due to differences in characteristics and an unexplained residual. The difference in mean wages can be written as:

$$\overline{W}_m - \overline{W}_f = (\overline{X}_m - \overline{X}_f)\hat{\beta}_m + (\hat{\beta}_m - \hat{\beta}_f)\overline{X}_f \equiv E + U, \tag{3}$$

where $\overline{W}_g$ and $\overline{X}_g$ denote the mean log wages and control characteristics of group g and $\hat{\beta}_g$ represents the estimated parameter from equation (2). While the first term stands for the effect of different productive characteristics (the endowment effect E), the second term represents the unexplained residual U which is due to differences in the estimated coefficients for both groups and is often referred to as discrimination effect.[3] Since the first use in the early 1970s, a number of authors have adopted and also extended the Blinder–Oaxaca approach.[4] For our meta-study, we accepted all estimates for log wage differentials, dummies, as well as the unexplained gender wage residual U and its derivatives. These estimates are taken as the dependent variable in our meta-regression analysis which we try to explain by the respective papers' data and method characteristics.

There has been much discussion about which variables to include in a wage regression that a Blinder–Oaxaca decomposition is based on.[5] Two questions are crucial. First, are the included characteristics affected by discrimination themselves? If yes, the estimated discrimination is understated. Second, do the included variables measure productivity comprehensively? If not, the estimate is biased

upwards or downwards. Consequently, skepticism exists whether U can be right-fully called discrimination effect – some argue the term unexplained residual may be more appropriate.[6] While an employer is assumed to have exact knowledge of all the relevant productive characteristics of an employee and can set the wage accordingly, the researcher usually possesses only the data for a restricted number of indicators for productivity. If the omitted variables correlate with sex, then U might capture not only discrimination, but unobserved group differences in productivity as well. In particular, it has been argued that less investment in on-the-job training, less experience, greater time in housework, and lower occupa-tional attainments of women may be voluntary choices made by women which are not adequately captured in the data and might be responsible for the 'unexplained residual' U.[7] Meta-analysis cannot tell us an ideal specification for an analysis of the gender wage gap. What it can do, however, is to give a critical assessment of how certain restrictions in a particular data set or the choice of a particular specification will affect the results.

4. Meta-Data

In order to make the data construction as transparent as possible, we used an easily accessible but universal research database. Following Stanley and Jarrell (1998), in November 2000, we searched the Economic Literature Index (EconLit) for any reference to: '(wage* or salar* or earning*) and (discrimination or differen*) and (sex or gender)'.[8] EconLit is the most comprehensive database for economic research papers. There is a bias toward internationally published research, which might be considered a welcome selection with regard to quality; on the other hand, non-English-language studies will be underrepresented, parti-cularly if they represent solely policy reports or unpublished papers from research institutes. However, correcting this bias seems impossible since there is no other suitable research database available.

Our EconLit search led to 1541 references of which a large fraction was theoreti-cal, or, in fact, covering an entirely different topic. The empirical papers were examined whether they actually used any regression analysis or simply reported mean ratios without holding productivity constant. Eventually, the desired estimates could be gained from 263 articles.[9] Some authors calculated the gender wage gap for several countries or time periods in one published paper. Likewise, they might use data from different distinct populations, like regional or sectoral entities. These estimates can be treated as independent estimates. Therefore, we divided the esti-mates from one paper into several 'studies' if the estimates have come from different time periods and/or different populations. This gives us 788 different studies.

Table 1 summarizes the distribution of our sample over time, where we coded a study for the 1980s, if its data related to the 1980s. The number of papers increased steadily over time, with a decreasing number in the 1990s, which is easily explained by a 'publication backlog' as well as a 'research backlog': data sets for the (late) 1990s are only available after some time.

Table 1. Data for Gender Wage Gaps.

	1960s	1970s	1980s	1990s	All
Number of papers	7	52	161	43	263
Number of different 'studies'	21	189	429	149	788
Number of different estimates	63	352	871	249	1535
Fraction of estimates					
USA	0.65	0.55	0.37	0.19	0.41
Europe	0.13	0.12	0.23	0.34	0.21
Other OECD	0.19	0.12	0.13	0.12	0.13
Post-communist countries	0	0	0.01	0.11	0.02
Africa	0	0.05	0.03	0.06	0.03
Asia	0	0.13	0.13	0.10	0.12
Latin-America	0.03	0.03	0.10	0.08	0.08
Mean total wage gap	0.51	0.43	0.30	0.26	0.33
Mean unexplained wage gap	0.23	0.22	0.20	0.19	0.20

Typically, authors present a number of estimates for each study, i.e. country and time unit. These estimates are usually based on different specifications of the regression model. Stanley and Jarrell (1998) selected only one estimate per paper for their meta-analysis. In particular, they chose 'the OLS estimate which the author seemed to promote as the best' (p. 955). We included all estimates the authors presented for a given study to avoid any possibility of a systematic bias when picking a certain estimate.[10] Furthermore, we wanted to make use of the information yielded by different estimates from the same data. For each study, all estimates as well as all the corresponding meta-independent variables, data characteristics, and methodology were collected and coded. (The meta-independent variables included in the analysis are listed in Table 2.) This procedure yields one observation in our meta-data set per reported estimate. In total, this gives us 1535 estimates of the gender wage gap, on average two estimates per study.[11]

However, there are two potential problems associated with allowing multiple estimates from one study: First, multiple estimates using the same data (same country and time period) are not independent from each other, leading to non-spherical error terms in the meta-regression. Second, there is the problem of biased sampling: if multiple estimates of one single study were treated as separate observations, studies with a larger number of estimates would receive more weight. We deal with these problems using a weighting scheme to correct for this bias (see Section 5.)

While Stanley and Jarrell (1998) use only US studies which are based on one of the broad national data sets (CPS, Census, or PSID),[12] we collected all estimates of the gender wage gap based on data for 64 countries. Table 1 also gives a regional breakdown of our data set. Whereas in the beginning of the sampling period, estimates for the US were in the majority, their share fell to merely 19% in the 1990s. Especially, in the later periods, a considerable amount of the estimates of the gender wage gap come from post-communist countries, Asia, Latin-America, or Africa.

Table 2. Meta-Independent Variables.

A) Paper		Mean	Standard deviation
Author female	percentage of authors who are female	0.28	0.36

B) Data sets			
New entries	1 if a study investigated the wages of new entrants only	0.02	0.13
Public sector	1 if a study investigated the wages of workers in the public sector only	0.09	0.29
Private sector	1 if a study investigated the wages of workers in the private sector only	0.12	0.32
Narrow occupation	1 if a study investigated the wages of workers of a narrowly defined occupation only	0.14	0.34
Low-prestige occupation	1 if a study investigated only low-prestige occupations (e.g. blue collar)	0.04	0.19
Medium-prestige occupation	1 if a study investigated only medium-prestige occupations (e.g. white collar)	0.07	0.25
High-prestige occupation	1 if a study investigated only high-prestige occupations (e.g. college graduates and academics)	0.18	0.38
Single only	1 if a study investigated only singles	0.04	0.20
Married only	1 if a study investigated only married people	0.03	0.17
Minority only	1 if a study investigated only minority or immigrant population	0.02	0.15
Majority only	1 if a study investigated only majority population	0.08	0.28
Source	0 if data come from administrative statistics 1 if data come from survey data	0.95	0.22
Full-time only	1 if a study included only full-time workers	0.32	0.47

C) Method of estimation			
Dummy variable	1 if a study used a dummy to investigate the gender wage gap and no Blinder–Oaxaca decomposition	0.22	0.41
IV	1 if a study used instrumental variables	0.01	0.10
Panel data	1 if a study used panel data	0.04	0.18
Heckman	1 if a study corrected for selectivity á la Heckman	0.24	0.42
Blinder–Oaxaca with female coefficients	1 if female wage structure was used for the decomposition instead of male one	0.21	0.41
Neumark	1 if Neumark decomposition was used	0.09	0.29
Cotton	1 if Cotton decomposition was used	0.01	0.11

| Brown | 1 if Brown et al decomposition was used | 0.01 | 0.11 |
| Reimers | 1 if Reimers decomposition was used | 0.01 | 0.09 |

D) Alternative measures of wages

No hourly wages	1 if a study used daily, monthly, or annual earnings	0.60	0.49
Hourly constructed	1 if a study used hourly wages computed from daily, weekly, monthly, or annual salary	0.16	0.37
Gross	0 if a study used net wages 1 if a study used gross wages	0.07	0.26

E) Variables for worker's characteristics

Potential experience	1 if a study used potential experience	0.50	0.50
Experience	1 if a study omitted worker's job experience	0.02	0.16
Race or immigrant	1 if a study failed to account for race or immigrant status	0.61	0.49
Marital status	1 if a study omitted worker's marital status	0.41	0.49
Kids	1 if a study omitted whether or not worker has children	0.71	0.46
Marital/kids interaction	1 if a study omitted interaction children * marital status	0.96	0.20
Training	1 if a study omitted on the job training	0.97	0.16
Tenure	1 if a study omitted tenure	0.73	0.44
Occupation	1 if a study omitted worker's occupation	0.55	0.50
Industry	1 if a study omitted worker's industry of employment	0.65	0.48
Government work	1 if a study omitted a government/private employment distinction	0.57	0.50
Union status	1 if a study omitted worker's union/nonunion status	0.75	0.43
Share of females in occupation	1 if a study omitted the percentage of women in the worker's job	0.88	0.33
FT-PT	1 if a study omitted worker's full-time/part-time status	0.51	0.50
Urban	1 if a study omitted SMSA, city size	0.63	0.48
Reg	1 if a study omitted worker's geographical area of employment	0.42	0.49
Working time	1 if a study omitted worker's working time	0.99	0.08

Figure 1 shows the development of the total wage gap (i.e. the raw differential in hourly log wages from the original data set) over time. The total wage gap falls significantly over time from around 65% ($e^{0.5} - 1$) in the 1960s to only 30% in the 1990s. The ratio of male to female wages declined by 0.8% per year during this period. Although the total wage differential has more than halved across our
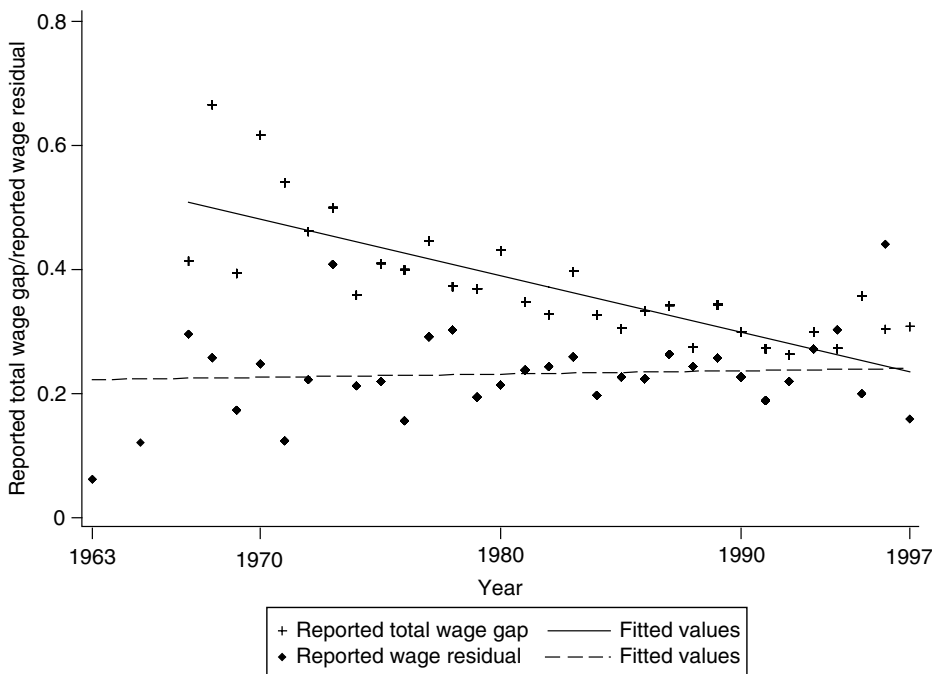
Figure 1. Reported log wage gaps over time.

time period 1963–1997, this decline is almost entirely due to an equalization of productive characteristics: females have become better educated and trained. The reported Blinder–Oaxaca wage residual is practically constant over time.

Figure 2 shows the reported total wage gap and the reported wage residual for the different countries, shown as simple averages over all estimates for the respective countries. (Country codes used in Figure 2 are given in Table 6.) In those countries plotted above the 45° line (e.g. Cote d'Ivoire, Tanzania, Korea, Kenya, Cyprus, Japan, Indonesia, and Nicaragua), women have lower endowments than men. Part of the total wage gap can therefore be attributed to differences in human capital. In those countries underneath the 45° line (e.g. Singapore, Guinea, Costa Rica, Sudan, Trinidad and Tobago, Philippines, and South Africa), the contrary is true. Women have higher endowments than men, nevertheless they are paid less. Considering their human capital, women, in fact, are more discriminated than suggested by the total wage gap.

5. Meta-Regression Analysis

Our meta-regression model takes the form:

$$R_j = \sum a_k Z_{kj} + b \ t_j + d \ c_j + \varepsilon_j \qquad j = 1, 2, ..., L, \qquad k = 1, 2, ..., M \qquad (4)$$

where R_j represents the unexplained log wage differential of study j, which can
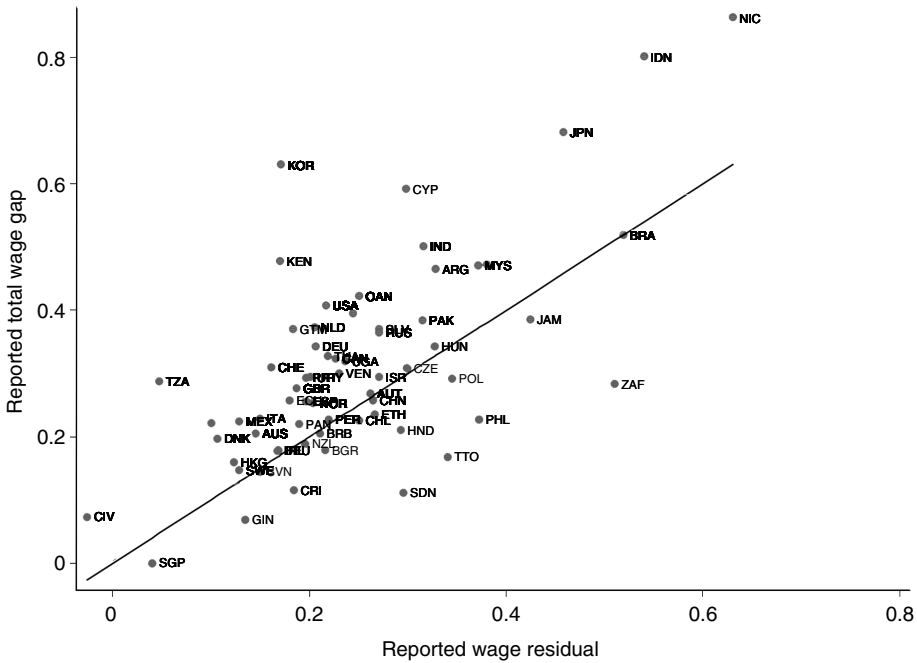
Figure 2. Reported log wage gaps for countries.

either be the Blinder–Oaxaca unexplained residual U_j from (3) or the coefficient of the gender dummy γ_j in (1), Z_{kj} are the k meta-independent variables, t_j and c_j are a set of time and country dummies, respectively; a_k, b, and d are parameters to estimate.

The meta-regressions presented in Table 3 include meta-independent variables describing the data set, the econometric technique and the type of wage information used, the inclusion of certain control variables in the original wage regressions and a dummy for the sex of the researcher. In addition, a full set of country and time dummies is included. The base category concerning the data set is always a random sample of the total population. Concerning the control variables, the base category is always the inclusion of the respective variable in the wage regressions. Critics often claim that in meta-analyses, apples and oranges may be mixed accidentally leading to artificial results (Furberg and Morgan, 1987). Calculated effects may differ between investigated studies, but there may also be no theoretical reason to expect a common parameter across them. Indeed, also gender wage differentials vary widely across countries and time periods. However, meta-regression analysis with data with different expected effects is less problematic, because time and country dummies can control these different effects. Furthermore, in Section 6.4., we relax our pooling assumption and present separate results for different regions.

Table 3. Meta-Regression Estimates.

Weighting scheme	(1) No weights	(2) Number of estimates in study	(3) (2) + journal rank	(4) (2) + precision of estimates	(5) (2) + square root of DF	(6) (2) + R^2 of wage regression
Author female	0.006 (0.013)	0.014 (0.015)	−0.033 (0.019)	−0.017 (0.015)	−0.006 (0.019)	0.030 (0.020)
New entrants	−0.103 (0.022)**	−0.087 (0.026)**	−0.187 (0.029)**	−0.077 (0.025)**	−0.087 (0.024)**	−0.104 (0.030)**
Full-time workers	0.013 (0.012)	0.013 (0.013)	0.052 (0.016)**	−0.007 (0.015)	−0.015 (0.018)	−0.003 (0.020)
Private sector	0.008 (0.026)	−0.005 (0.025)	−0.041 (0.024)	0.025 (0.028)	−0.003 (0.023)	0.013 (0.032)
Public sector	−0.049 (0.020)*	−0.067 (0.020)**	−0.080 (0.028)**	−0.061 (0.033)	−0.061 (0.022)**	−0.037 (0.028)
Narrow occupation	−0.051 (0.018)**	−0.050 (0.018)**	−0.035 (0.026)	−0.026 (0.025)	−0.058 (0.022)**	−0.034 (0.024)
Low-prestige occupation	0.048 (0.020)*	0.049 (0.019)*	0.144 (0.018)**	−0.029 (0.036)	0.092 (0.022)**	−0.032 (0.037)
Medium-prestige occupation	−0.046 (0.018)*	−0.041 (0.016)*	−0.056 (0.016)**	−0.072 (0.037)	−0.051 (0.021)*	−0.073 (0.031)*
High-prestige occupation	−0.126 (0.016)**	−0.121 (0.015)**	−0.131 (0.011)**	−0.098 (0.027)**	−0.154 (0.015)**	−0.086 (0.027)**
Singles	−0.139 (0.023)**	−0.133 (0.019)**	−0.119 (0.023)**	−0.119 (0.043)**	−0.183 (0.040)**	−0.147 (0.028)**
Married	0.091 (0.028)**	0.076 (0.024)**	0.105 (0.022)**	0.067 (0.038)	0.029 (0.053)	0.131 (0.047)**
Minority	−0.046 (0.019)*	−0.073 (0.021)**	−0.013 (0.030)	0.011 (0.024)	−0.074 (0.017)**	−0.103 (0.028)**

	(1)	(2)	(3)	(4)	(5)	(6)
Blinder–Oaxaca with female coefficient	− 0.016 (0.008)	− 0.017 (0.009)	− 0.018 (0.011)	− 0.003 (0.003)	− 0.010 (0.008)	0.001 (0.012)
Neumark decomposition	0.019 (0.009)*	0.002 (0.013)	0.048 (0.012)**	0.017 (0.009)	− 0.017 (0.017)	− 0.010 (0.015)
Reimers decomposition	− 0.002 (0.038)	0.019 (0.024)	0.061 (0.035)	0.019 (0.018)	0.090 (0.025)**	0.002 (0.048)
Cotton decomposition	− 0.008 (0.022)	− 0.017 (0.030)	0.039 (0.020)	0.011 (0.017)	− 0.016 (0.033)	− 0.004 (0.028)
Brown decomposition	− 0.020 (0.018)	− 0.025 (0.019)	− 0.003 (0.045)	0.024 (0.021)	− 0.020 (0.019)	0.033 (0.038)
Dummy variable	0.011 (0.013)	0.010 (0.014)	0.016 (0.017)	0.059 (0.020)**	0.021 (0.018)	0.007 (0.019)
IV	− 0.027 (0.023)	− 0.041 (0.028)	− 0.045 (0.034)	− 0.062 (0.026)*	− 0.045 (0.028)	− 0.031 (0.040)
Panel data	− 0.004 (0.025)	0.015 (0.027)	0.042 (0.033)	0.091 (0.067)	− 0.052 (0.041)	0.143 (0.036)**
Heckman selection	− 0.005 (0.010)	0.002 (0.014)	− 0.016 (0.013)	0.019 (0.010)	− 0.031 (0.015)*	0.024 (0.019)
No hourly wages	0.031 (0.016)*	0.025 (0.019)	0.038 (0.024)	0.040 (0.023)	0.026 (0.019)	− 0.004 (0.030)
Hourly wages constructed	− 0.001 (0.017)	− 0.020 (0.020)	− 0.009 (0.022)	0.053 (0.018)**	− 0.029 (0.022)	− 0.038 (0.026)
Gross wages	− 0.019 (0.033)	− 0.021 (0.030)	− 0.065 (0.045)	− 0.015 (0.067)	− 0.105 (0.052)*	− 0.084 (0.061)
Potential experience	0.001 (0.012)	0.021 (0.014)	0.029 (0.015)	− 0.041 (0.018)*	0.008 (0.016)	0.033 (0.016)*

(*continued*)

Table 3. *Continued.*

Weighting scheme Variables missing in wage regression	(1) No weights	(2) Number of estimates in study	(3) (2) + journal rank	(4) (2) + precision of estimates	(5) (2) + square root of DF	(6) (2) + R^2 of wage regression
Experience	− 0.002 (0.025)	− 0.022 (0.026)	− 0.044 (0.026)	− 0.063 (0.029)*	− 0.036 (0.030)	− 0.015 (0.032)
Race or immigrant	0.003 (0.014)	0.008 (0.017)	0.039 (0.017)*	0.008 (0.021)	0.019 (0.017)	− 0.026 (0.022)
Marital status	− 0.028 (0.012)*	− 0.038 (0.014)**	− 0.010 (0.018)	0.037 (0.020)	− 0.030 (0.017)	− 0.036 (0.018)*
Kids	0.019 (0.014)	0.027 (0.016)	− 0.017 (0.019)	− 0.065 (0.018)**	0.028 (0.018)	0.040 (0.020)*
Marital/kids interaction	− 0.071 (0.038)	− 0.024 (0.043)	0.005 (0.040)	− 0.014 (0.050)	− 0.089 (0.040)*	− 0.001 (0.063)
Training	− 0.036 (0.024)	− 0.048 (0.028)	− 0.022 (0.010)*	− 0.022 (0.021)	− 0.063 (0.023)**	− 0.069 (0.031)*
Tenure	0.048 (0.010)**	0.047 (0.012)**	0.040 (0.016)*	0.077 (0.016)**	0.058 (0.014)**	0.047 (0.014)**
Occupation	0.002 (0.010)	− 0.002 (0.011)	− 0.010 (0.016)	0.030 (0.014)*	0.012 (0.011)	0.019 (0.014)
Industry	0.015 (0.013)	0.020 (0.013)	0.024 (0.018)	− 0.011 (0.018)	0.013 (0.014)	− 0.012 (0.019)
Government work	0.007 (0.016)	− 0.005 (0.016)	0.025 (0.017)	− 0.036 (0.017)*	0.017 (0.018)	− 0.019 (0.020)
Union status	0.022 (0.012)	0.020 (0.015)	0.042 (0.019)*	0.035 (0.018)	0.045 (0.016)**	0.024 (0.022)

Share of females in occupation	0.073 (0.012)**	0.074 (0.015)**	0.068 (0.009)**	0.053 (0.013)**	0.083 (0.015)**	0.092 (0.026)**
Full time/part time	− 0.000 (0.012)	0.014 (0.013)	0.023 (0.015)	− 0.022 (0.013)	− 0.010 (0.015)	0.012 (0.019)
Observations	1527	1527	1527	1067	1221	907
Adjusted R^2	0.52	0.52	0.78	0.81	0.66	0.52

Note: Other variables in the regressions include: indicators whether the sample was official data or survey data, whether gender wage differentials were the main topic of the paper, dummies for a data set with only workers from the majority population in the sample; dummies for regional and urban status missing; dummies if information in the paper about measures of wages, the used data set and the gender of the researcher was unknown. Moreover, all regressions include a full set of country and time dummies. Robust standard errors are expressed in parentheses. * and ** indicate significance at 5% and 1% level, respectively.

5.1 *Econometric considerations*

As has been mentioned, there are multiple estimates available for each study. Consequently, we had to adopt a particular weighting scheme to deal with the lack of independence of these estimates. Column (1) in Table 3 presents the results for the unweighted estimates. From column (2) on, however, all estimates of one study (same country and time period; i.e. same data set) are weighted with the inverse of the number of estimates contained in one study. Moreover, a clustering approach is used in all specifications to correct for a possible upward bias in the precision of the estimates due to non-spherical standard errors, caused by the fact that some observations come from the same study. This robust variance estimate adjusts for within-cluster correlation, where in this case the clusters are defined as the different studies (Froot, 1989).

A further problem of meta-regression analysis concerns the quality of the study. Meta-analysis treats all studies alike. This is not always fortunate, because the researcher might have some priors about how a good study should look like. Meta-studies typically tackle the question of 'study quality' indirectly by including quality characteristics as a part of meta-independent variables, thus showing their effect on the dependent variable. For instance, a meta-study might estimate the effect of a more advanced econometric technique on a regression coefficient. Another approach, however, would be to weight well-done studies more heavily than others. We, therefore, experimented with different weighting schemes in columns (3) – (6), always in addition to the weighting that was already applied in column (2).

A usual approach in meta-analysis is to take the precision of the estimate (in general the inverse of the standard error) as a quality indicator. This cannot be done in our case, because, in general, the users of the Blinder–Oaxaca decomposition do not report the precision of this constructed indicator.[13] Therefore, at first, we used only studies published in journals and applied the citation-based journal rankings from Laband and Piette (1994) as weights. This scheme is agnostic about our own priors of study quality, but assumes that the peer-review process does a good job in letting the very reliable studies be published in the best journals. A drawback of this approach is that studies from exotic countries often find it much harder to get access to top-notch international journals. The next scheme, applied in column (4), uses only those papers reporting more than one estimate per study. One could argue that if a researcher used different specifications and got the same results, his/her study should be judged as more reliable. Therefore, we weight with a precision index of the estimates, i.e. with the inverse of the coefficient of variation among the estimates within one study. Of course, this weighting scheme treats the different estimates within a study alike, which might not be appropriate when the researcher wants to contrast different methodological approaches and single out the best one. Sampling theory would suggest that the absolute value of the *t*-statistic would be proportional to the square root of the degrees of freedom in the regression. Although we do not have *t*-statistics – because the wage differential is calculated based on the male and

female wage regressions – we can still use the square root of the degrees of freedom as a weighting scheme in column (5). Finally, column (6) uses the weighted mean of the R^2 of the original male and female wage regressions as a weight for the precision in the calculation of the gender wage gap.

Publication bias occurs when journal editors tend to publish papers with significant results only (Ashenfelter *et al.* 1999). It can seriously harm meta-regression analysis when studies with low or insignificant results are systematically missing, because the numerical size of the effect will be overestimated. However, for the unexplained residual of gender wage gap this is less of a problem, since typically no standard errors are reported. While this might be unfortunate for the general quality of this research stream, it reduces possibilities for publication bias considerably, because both researchers and editors cannot judge a paper according to the statistical significance of the result. Also the sign of the gender wage gap should not cause a 'file drawer problem', as Stanley and Jarrell (1998) have pointed out. Since most researchers accept the fact that gender wage differentials exist, a study finding no or reverse discrimination 'is more, not less, likely to be published' (Stanley and Jarrell, 1998, p. 954). Stanley (2005) calls the fact that strong effects, no matter in which direction, are more likely to be published 'type II selection' and concludes 'Fortunately, if there is type II selection without any noticeable directional selection, it is unlikely to materially affect the overall assessment of either conventional narrative reviews or meta analyses' (Stanley, 2005).

Another general problem in meta-regression analysis is the question whether the usual asymptotic assumptions for the error term in the regression are fulfilled. The first reason for concern is the fact that the dependent variable is a constructed variable based on original microdata. The usual solution for constructed regressors (Murphy and Topel, 1985) is not applicable in our case, because the statistical precision of the calculated gender wage gap is unknown. The second issue concerns correct sampling. What is the appropriate population to sample our data points from? One possibility is the population of all existing countries during the time period 1960–2000, the other possibility is the population of studies on gender wage gaps in these countries in the given time period. We are reasonably optimistic to have a random sample of existing studies, with possibly a bias in favor of English-language literature; but we have to be less optimistic to have a random sample of gender wage gaps for each country. Moreover, some of the existing studies of different authors might have used the same or very similar data but different methods, which raises concerns for non-independence of data points. There is no clear solution for this; neither a fully convincing correction for the constructed-regressor problem nor for the unclear sampling scheme can be offered. We have to take these drawbacks of meta-regression analysis into account and have to interpret our results with appropriate caution.[14] We will, therefore, place particular emphasis on robustness of our results, i.e. consistencies in coefficients across different specifications.

6. Results

6.1 *Effects of Data and Method*

Although all of the above-described weighting approaches are somewhat arbitrary and have some particular drawbacks, the general results are very similar.[15] The biggest – and very consistent – impact on the gender wage gap results from the type of data set used. In comparison to a random sample of the population, the gender wage gap is much lower if only a sample of new entrants in the labor market is investigated. Likewise, the wage gap is lower in the public sector and if only a narrow occupation is studied, because in the latter case, holding productivity equal is much easier. Interestingly, the wage gap is higher in the sample with only low-prestige occupations (blue-collar jobs) and lower for only high-prestige jobs (e.g. college graduates and academic jobs) as compared to a sample including all occupations. In accordance with Becker's household specialization model (1991), the wage gap is highest for married employees and significantly lower for singles. Among minority workers, the gender wage gap is somewhat smaller.[16]

The impact of other variables is less consistent across specifications. In terms of decomposition methods, it does not matter much whether the authors used only a dummy variable approach or one of the variants of the Blinder–Oaxaca decomposition technique; we get some significant coefficients, but no consistent picture across specifications. Instrumental variables approaches – which, in general, instrument for the endogeneity of work experience and/or training – result in considerably lower gaps; they are only marginally significant, because IV estimates occur only in 1% of all cases. The use of panel data and sample-selection techniques à la Heckman does not seem to matter in a consistent way. The income measure in original microdata is usually given by monthly earnings or hourly wages. One would expect that hourly wages lead to lower wage differentials than other measures, because women often work fewer hours and have more work interruptions, which are typically not observable in the data. However, in this model, we do not find a significant effect whether hourly wages or monthly (annual or weekly) incomes are given in the original data. The variable 'potential experience' captures whether in the original data 'work experience' was not explicitly given, but instead calculated as age-6-years of education. One would expect a higher unexplained gender wage gap if potential experience was used in the wage regression, again due to women's more frequent career interruptions. However, this is not consistently reflected in our results. Next, we consider the specification of the wage regressions. What effect does the inclusion or exclusion of a particular variable have on the estimated wage gap? As has been noted before, estimates on the gender wage gap can be biased for two reasons: (i) some productive characteristics are observed by the firm, but not by the econometrician. In most cases, this will lead to an upward bias in the resulting gender wage gap or discrimination component; (ii) some of the control variables might themselves be caused by unequal treatment of the sexes (e.g. occupational choice and promotion). Inclusion of such variables might give rise to a downward bias,

because possible discrimination in promotion or occupational choice is falsely regarded as a difference in productive characteristics. In general, this reasoning could be valid for most of the usual control variables, e.g. job tenure or work experience. To use a consistent specification, we include indicators for the absence of each of these variables in the respective papers, while the base category is the inclusion.[17] That is, the variable 'marital status = 1' indicates that the author of a paper neglected the marital status of the individuals studied in his wage regression.

Also, the impact of these variables on the gender wage gap is much lower – and less consistent – than the effect of the sample restrictions: Missing marital status as well as missing training in the wage regression has a negative effect on the wage gap, whereas missing tenure has a positive effect. The marital status of an individual could be interpreted as a productivity indicator. Household responsibilities may make married females less productive at the job, while males benefit from their wife's reproductive work, become more productive, and earn a marriage premium. If a researcher neglects this productivity indicator in the wage regression s(he) erroneously calculates a downward biased gender wage gap. However, married men may also simply receive preferential treatment. Hersch and Stratton (2000) find that married and single men spend virtually the same time on household tasks which conflicts with the specialization theory. In that case including marriage status may cause an upward bias. As tenure is an important productivity component and females often have less tenure, neglecting tenure in a wage regression can lead to a serious overestimation of the discrimination component. Missing union status has a consistently positive effect on the gender wage gap, because union jobs tend to be better-paid, male-dominated jobs. Also, if information about the share of females in the respective occupation is missing, this increases the gender wage gap. There are two possible reasons for this outcome. Either occupational choice is governed by preferences and wages correctly reflect productivity or pre-market discrimination in schooling as well as discrimination in hiring leads to occupational crowding. If the second is true, including a variable on the female domination of a job produces a downward bias of the measured discrimination.

Interestingly, in contrast to Stanley and Jarrell (1998 and 2004), we find no consistent impact of the gender of the researcher on the outcome of the study. One might suspect that women experience discrimination more frequently on a personal basis and therefore are more susceptible to accept higher estimates of gender wage gaps; however, Stanley and Jarrell, in fact, found the opposite effect. They interpret their finding that women compensate for potential bias due to their gender membership. Considering the results from our data, this does not seem to be the case internationally. Only in the journal-rank-weighted specification, the wage gap is somewhat lower if the researcher was a female. One could bravely interpret this finding in such a way that women have to be relatively more prudent if they want to get access to top economics journals.

What are the relative contribution of data selection and the choice of econometric methods in the explanation of the variance in gender wage gaps? To

answer this question, we ran separate OLS regressions, in the one case including only the 19 data-selection variables, in the other only the 24 method variables (without country, time, and gender of researcher dummies). The resulting R^2 are presented in Table 4 and confirm the view that the choice of data set is quantitatively more important than the choice of method. Whereas for the entire time period around 20% of the variance in gender wage gaps is explained by the choice of data, the choice of econometric method explains only 12%. For each decade, the R^2 for methods are higher than for the entire time period and the difference between data and methods seems to be minor. However, this is due to the fact that the coefficients for data characteristics are more stable for each time period and therefore less affected from a pooling of the data over decades. Consequently, the data characteristics have the most consistent effect on the calculated gender wage gap.

6.2 *Fixed-Effects Estimates*

The most time-consuming task of meta-analysis is to carefully read and code all the details of the analyzed papers. The coding of method and data used can only be as precise as the description in the papers provided by the authors. The accuracy of the coding also depends on how well the examined features can be quantified. Some features of a research paper, e.g. specificities of the data set, the exact wording in the underlying questionnaire, how the researcher is treating the raw data, and minor econometric decisions coming up in the course of the research, may remain unknown. Therefore, a panel data (fixed-effects) approach might offer a useful tool to control for these paper-specific effects, which are unobservable to the meta-econometrician. This is possible, because several observations sharing these same study or paper characteristics are available in the primary data set. There are two possibilities for the meta-analyst: (i) take the paper as the unit of observation and treat all estimates within a paper as deviations from the paper's mean; (ii) take the study (i.e. one country and time period within a paper) as the unit of observation. Table 5 reports fixed-effects estimates for both of these variants. It has to be noted, though, that the coefficients in these fixed-effects models are identified only by papers (or studies) having several estimates. Therefore, the precision of some of the coefficients must suffer due to

Table 4. Are Data Selection or Econometric Methods More Important in Explaining the Variance in Gender Wage Gaps?

	1960s	1970s	1980s	1990s	All
Contribution of data selection					
R^2	0.62	0.30	0.22	0.26	0.20
R^2 adjusted	0.55	0.27	0.20	0.21	0.19
Contribution of econometric methods					
R^2	0.68	0.26	0.13	0.25	0.12
R^2 adjusted	0.56	0.20	0.11	0.18	0.11

Table 5. Panel Model (Fixed Effects).

Group indicator for fixed effect	Paper	Paper	'Study' within paper
New entrants	− 0.093 (0.037)*	− 0.093 (0.036)*	− 0.091 (0.042)*
Full-time workers	− 0.046 (0.037)	− 0.047 (0.036)	− 0.043 (0.054)
Private sector	0.029 (0.025)	0.040 (0.025)	
Public sector	− 0.024 (0.025)	− 0.008 (0.026)	− 0.011 (0.047)
Narrow occupation	0.021 (0.024)	0.022 (0.023)	
Low-prestige occupation	0.074 (0.018)**	0.073 (0.017)**	− 0.168 (0.104)
Medium-prestige occupation	− 0.020 (0.017)	− 0.025 (0.017)	
High-prestige occupation	− 0.079 (0.017)**	− 0.081 (0.017)**	
Singles	− 0.197 (0.022)**	− 0.201 (0.022)**	− 0.303 (0.066)**
Married	0.085 (0.025)**	0.080 (0.024)**	
Minority	− 0.129 (0.041)**	− 0.129 (0.040)**	− 0.086 (0.099)
Majority	− 0.071 (0.043)	− 0.071 (0.042)	
Blinder–Oaxaca with female coefficient	− 0.011 (0.007)	− 0.011 (0.007)	− 0.010 (0.006)
Neumark decomposition	0.024 (0.011)*	0.025 (0.011)*	0.027 (0.009)**
Reimers decomposition	− 0.027 (0.033)	− 0.027 (0.032)	− 0.027 (0.027)
Cotton decomposition	− 0.001 (0.027)	− 0.001 (0.026)	− 0.001 (0.022)
Brown decomposition	− 0.006 (0.029)	− 0.008 (0.028)	0.006 (0.045)
Dummy variable	0.040 (0.033)	0.042 (0.032)	− 0.005 (0.033)
IV	0.021 (0.034)	0.022 (0.034)	0.007 (0.031)
Panel data	− 0.127 (0.042)**	− 0.092 (0.042)*	− 0.226 (0.058)**
Heckman selection	− 0.013 (0.009)	− 0.012 (0.008)	− 0.019 (0.008)*

Table 5. *Continued.*

Group indicator for fixed effect	Paper	Paper	'Study' within paper
No hourly wages	0.103 (0.038)**	0.101 (0.037)**	0.102 (0.033)**
Gross wages	0.033 (0.027)	− 0.000 (0.083)	
Potential experience	0.042 (0.024)	0.027 (0.024)	0.033 (0.023)
Variables missing in wage regression			
Experience	0.062 (0.032)	0.077 (0.031)*	0.104 (0.037)**
Race or immigrant	0.280 (0.028)**	0.038 (0.061)	0.083 (0.066)
Training	− 0.013 (0.029)	− 0.016 (0.028)	− 0.003 (0.025)
Tenure	0.026 (0.019)	0.032 (0.019)	0.039 (0.023)
Occupation	0.024 (0.013)	0.026 (0.013)*	0.037 (0.014)*
Industry	0.026 (0.016)	0.020 (0.015)	0.022 (0.016)
Government work	0.001 (0.022)	0.008 (0.022)	0.046 (0.026)
Union status	0.002 (0.025)	0.022 (0.029)	0.018 (0.028)
Share of females in occupation	0.056 (0.017)**	0.056 (0.017)**	0.054 (0.014)**
Full time/part time	0.029 (0.028)	0.030 (0.028)	0.033 (0.028)
Urban	− 0.040 (0.048)	0.034 (0.064)	− 0.028 (0.111)
Region	− 0.046 (0.025)	− 0.051 (0.024)*	− 0.070 (0.024)**
Constant	− 0.061 (0.064)	0.067 (0.125)	− 0.004 (0.077)
Year dummies	No	Yes	No
Country dummies	No	Yes	No
Observations	1527	1527	1527
Number of groups	262	262	775
R^2 within	0.28	0.35	0.17

Note: Standard errors are expressed in parentheses. * and ** indicate significance at a 5% and 1% level, respectively.

low variation within the group. Regardless of the unit of the fixed-effect the results are rather robust; this applies also in comparison to the OLS results from Table 3.

Note that the terms 'fixed' and 'random effects' have different meanings in the econometrics and meta-analysis literature. We use the econometric terms which refer to methods in panel data analysis, in particular to the way how the error term in a panel data model is specified – either containing a unit-specific fixed part or behaving like a random variable. In the meta-analysis literature, fixed and random effects relate to the weights in the meta-analysis. In the fixed-effects specification, each estimate is assigned a weight inversely proportional to its variance. In the random-effects method, it is assumed that the studies are a random sample from a larger population of studies, and that the population effect sizes are randomly distributed about a population mean. In econometric terms this would be called a random-coefficient model (Abreu *et al.*, 2005).

Again, sample restrictions turn out to be very important; if the sample includes only new entrants, single workers or high-prestige occupations, wage differentials are lower, likewise if the sample is ethnically homogeneous. In contrast to the OLS regressions, the effects of econometric methods come out more explicitly. Estimates using panel methods or sample-selection techniques find lower wage gaps; estimates using the Neumark decomposition technique as compared to the Blinder–Oaxaca approach find higher wage gaps. While previous regression results did not show any systematic effects for the unit of wage measure available in the data, the fixed-effects model indicates that the use of non-hourly wages (in general monthly or yearly incomes) results in significantly higher gender wage gaps as would be expected. A similar situation arises, if experience was not explicitly included in the wage regression: measured unexplained wage gaps are considerably overestimated in such a case.[18] The panel model also demonstrates more clearly that controlling for occupation decreases the gender wage gap which is no surprise considering that women more often work in low-paid jobs.

6.3 *Pattern Across Countries and Time*

To report and assess a pattern across countries and time, it is useful to synthesize the data in a particular way to eliminate contaminating effects of different methods and data sets. In this step, we wonder how wage gaps would look like if all the authors had used data with identical characteristics and applied identical methods. To investigate this, we calculate a 'meta wage residual' which is what authors would have received if they had all used the same, rather conservative, design: only single individuals from an otherwise representative population would have been considered, all control variables would be included and sample-selection procedures would be applied as well as an instrumental variables approach to control for endogeneity of human capital variables. Practically, such an approach leads to the lowest gender wage gap empirically obtainable. Of course, our chosen design is only one – and in a way an arbitrary one – of a large number of possibilities. Given the linear OLS regression we use, other

choices would simply shift the line in Figure 3 representing the gender wage residual up or down, but would leave the slope unchanged. At this point, we are only interested in an interpretation of the time (and country) effects; therefore we use a weighting scheme, which weights by the number of observations in the meta-regression per year and country.

Figure 3 illustrates the time trend of the reported wage residual (i.e. the Blinder–Oaxaca wage gap from the examined papers[19]) and the 'meta wage gap'.[20] While the reported wage residual shows a slight upward trend over time, our constructed meta wage residual (in logs) falls with a rate of −0.0017 per year. This means that the ratio of what women would earn absent of discrimination to their actual wages decreases by only 0.17% annually. This is a rather moderate improvement over time.[21] This discrepancy between the development of the reported wage residual and our 'ideal' meta wage residual could be explained by a different choice of data sets over time, which might have led researchers in the early years to a relatively low 'discrimination' component.[22] Looking in detail at trends for country groups, we find that the meta wage residual for the US dropped by 0.003 per year, whereas for other OECD countries (Canada, Australia, and New Zealand) the gap dropped by 0.008 per year. This means that the ratio of what women should earn absent of discrimination and their actual earnings decreased by 0.3% annually in the US and 0.8% in the other
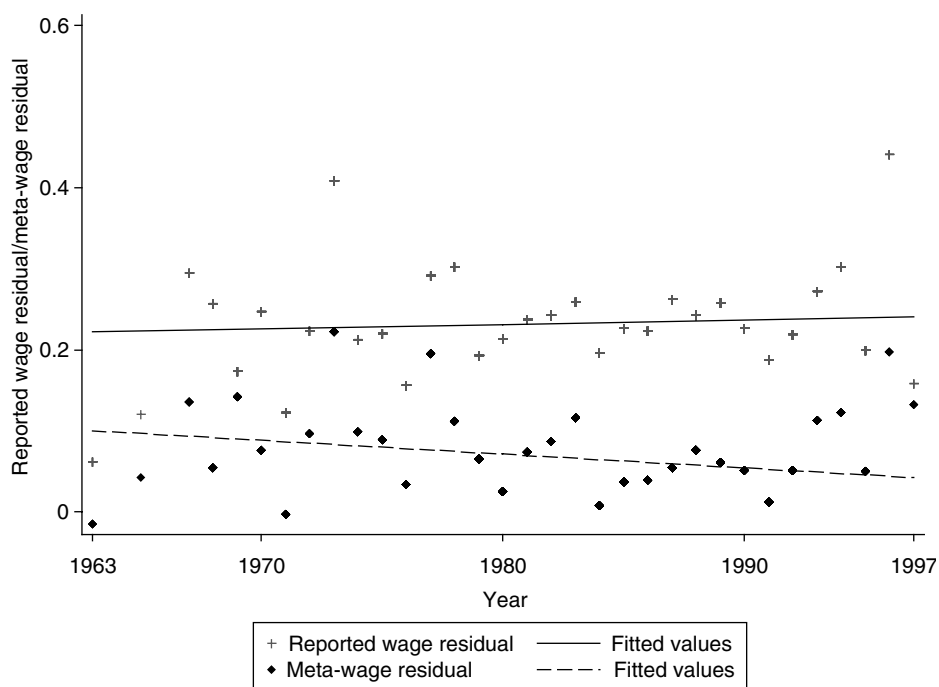

Figure 3. Wage residuals over time.

OECD countries. For Europe, we find a smaller, statistically insignificant, reduction of 0.2% per year. The largest decline was observed for post-communist countries with 1.9% per year – where we basically have observations only for the last few years. There was no discernible trend for Africa and Latin America – for Asia, the trend was even positive (0.4% per year). Stanley and Jarrell (1998, p. 966) calculated a drop in their meta-wage residual of more than 1% per year for the US and later (2004) revised this estimate to 0.6% per year, which is still a larger decline as our data would suggest for the US. Table 6 reports the coefficients for the respective country dummies from our main specification (Table 3/(2), weighting by the number of estimates in the study), together with the applied country codes and the number of observations we have for each country. These estimates indicate a particularly low meta-wage residual for countries like Barbados, Colombia, Cote d'Ivoire, Hong Kong, Italy, Kenya, Mexico, New Zealand, Slovenia, and Tanzania, while the wage residual in Brazil, Indonesia, and Nicaragua is especially high.

6.4 Estimates for Different Regions

So far, we have only included country dummies in our meta-regression to account for cultural differences in the gender wage gap. However, not only the level of the gender wage differential may be regionally different, there may also be an interaction effect with certain variables. For example, there could be cultural differences with respect to discrimination concerning new entries in the labor market. In one culture, new entrants could face higher gender wage differentials, because women have to prove themselves before they are accepted as good workers. In another culture, however, young workers, males and females, who come straight from school may be regarded as equally productive, in contrast to older workers. Older men may be expected to have invested more in their unobservables than women, which leads to higher wage differentials for more mature workers.

For this reason, we split our sample into four different categories: USA, Europe, other OECD countries, and the rest of the World (ROW, which basically covers all the developing and the former communist countries), and run our meta-regression for these regions separately while still including country dummies.[23] The results are presented in Table 7.[24] Wald tests have been performed to test whether coefficients are equal across regions. The gender wage gap is generally lower in the public sector as well as in high-prestige occupations worldwide (although not always significantly so). While the gender wage gap is always significantly lower if only single individuals are investigated, it tends to be higher if only married people are examined. If only one ethnic group, minority or majority, is examined, this also tends to decrease the gender wage gap in all regions. Strong differences in the results for the different regional areas, however, are observed concerning the effects of different methods: use of Neumark decomposition or dummy variables. While the use of panel data has not led to a significant effect in the joint analysis, for Europe we do find the expected effect. If unmeasurable personal fixed effects are taken into account, by the use of panel

Table 6. Country Effects from the Meta-Regression.

Abbreviation	Country	Coefficient	n
ARG	Argentina	Base	8
AUS	Australia	− 0.135 (0.039)**	63
AUT	Austria	− 0.084 (0.043)	28
BRB	Barbados	− 0.247 (0.084)**	4
BOL	Bolivia	0.004 (0.037)	2
BRA	Brazil	0.134 (0.061)*	19
BGR	Bulgaria	− 0.154 (0.035)**	1
CAN	Canada	− 0.104 (0.036)**	60
CHL	Chile	− 0.052 (0.052)	14
CHN	China	− 0.088 (0.041)*	11
COL	Colombia	− 0.246 (0.060)**	6
CRI	Costa Rica	− 0.182 (0.044)**	8
CIV	Cote d'Ivoire	− 0.359 (0.093)**	9
CYP	Cyprus	− 0.015 (0.062)	2
CZE	Czech Republic	− 0.124 (0.042)**	1
DNK	Denmark	− 0.149 (0.053)**	20
DEU	East Germany	− 0.194 (0.070)**	5
ECU	Ecuador	− 0.098 (0.047)*	2
SLV	El Salvador	− 0.097 (0.045)*	6
ETH	Ethiopia	− 0.013 (0.064)	9
GTM	Guatemala	− 0.178 (0.035)**	2
GIN	Guinea	− 0.164 (0.057)**	2

HND	Honduras	− 0.053 (0.032)	1
HKG	Hong Kong	− 0.202 (0.050)**	10
HUN	Hungary	0.028 (0.044)	4
IND	India	0.000 (0.046)	23
IDN	Indonesia	0.235 (0.067)**	8
IRL	Ireland	− 0.136 (0.038)**	22
ISR	Israel	− 0.010 (0.043)	11
ITA	Italy	− 0.234 (0.054)**	13
JAM	Jamaica	0.066 (0.160)	3
JPN	Japan	0.031 (0.057)	14
KEN	Kenya	− 0.225 (0.062)**	6
MYS	Malaysia	− 0.054 (0.057)	19
MEX	Mexico	− 0.205 (0.034)**	22
NLD	Netherlands	− 0.168 (0.054)**	15
NZL	New Zealand	− 0.205 (0.039)**	1
NIC	Nicaragua	0.301 (0.052)**	6
NOR	Norway	− 0.108 (0.039)**	54
PAK	Pakistan	− 0.059 (0.069)	21
PAN	Panama	− 0.086 (0.044)*	2
PER	Peru	− 0.050 (0.043)	22
PHL	Philippines	0.007 (0.049)	4
POL	Poland	− 0.078 (0.042)	1
PRT	Portugal	− 0.121 (0.066)	10

(*continued*)

Table 6. *Continued.*

Abbreviation	Country	Coefficient	*n*
RUS	Russia	− 0.146 (0.054)**	21
SGP	Singapore	− 0.098 (0.064)	5
SVN	Slovenia	− 0.220 (0.035)**	1
ZAF	South Africa	0.038 (0.105)	2
KOR	South Korea	− 0.074 (0.046)	38
ESP	Spain	− 0.128 (0.043)**	13
SDN	Sudan	0.059 (0.133)	3
SWE	Sweden	− 0.174 (0.041)**	17
CHE	Switzerland	− 0.123 (0.042)**	16
OAN	Taiwan	− 0.016 (0.047)	77
TZA	Tanzania	− 0.237 (0.047)**	13
THA	Thailand	− 0.104 (0.052)*	4
TTO	Trinidad&Tobago	− 0.168 (0.110)	2
GBR	UK	− 0.079 (0.038)*	87
USA	USA	− 0.061 (0.032)	614
UGA	Uganda	− 0.071 (0.056)	9
URY	Uruguay	− 0.169 (0.036)**	8
VEN	Venezuela	− 0.087 (0.047)	4
DEU	(West) Germany	− 0.117 (0.040)**	19

Note: These coefficients are from the meta-regression in Table 3, Column (2).

Table 7. Results Across Regions.

	USA	Europe	Other OECD	ROW	Wald test (p)
Author female	− 0.051 (0.019)**	0.070 (0.029)*	− 0.038 (0.055)	0.151 (0.068)*	0.000
New entrants	− 0.110 (0.037)**	− 0.056 (0.042)			0.000
Full-time workers	− 0.003 (0.026)	− 0.051 (0.028)	0.059 (0.045)	− 0.103 (0.068)	0.049
Private sector	0.001 (0.027)	− 0.075 (0.037)*	− 0.036 (0.065)	0.034 (0.085)	0.298
Public sector	− 0.063 (0.028)*	− 0.147 (0.043)**	− 0.080 (0.068)	− 0.005 (0.073)	0.244
Narrow occupation	− 0.054 (0.025)*	0.012 (0.036)	0.029 (0.041)	− 0.790 (0.118)**	0.000
Low-prestige occupation	− 0.015 (0.034)	− 0.042 (0.038)	0.038 (0.030)	0.173 (0.017)**	0.000
Medium-prestige occupation	− 0.101 (0.035)**	− 0.035 (0.029)	− 0.007 (0.011)	− 0.050 (0.016)**	0.002
High-prestige occupation	− 0.148 (0.029)**	− 0.023 (0.026)	− 0.154 (0.031)**	− 0.098 (0.012)**	0.000
Singles	− 0.112 (0.040)**	− 0.099 (0.019)**	− 0.086 (0.035)*	− 0.404 (0.073)**	0.001
Married	0.142 (0.073)	0.075 (0.023)**	0.059 (0.032)	0.006 (0.062)	0.406
Minority	− 0.040 (0.024)	− 0.148 (0.098)	− 0.226 (0.062)**		0.000
Majority	0.028 (0.023)	− 0.096 (0.095)	− 0.123 (0.062)*	− 0.401 (0.054)**	0.000
Blinder–Oaxaca with female coefficient	− 0.011 (0.013)	− 0.013 (0.014)	− 0.022 (0.031)	− 0.006 (0.016)	0.952
Neumark decomposition	0.062 (0.014)**	− 0.075 (0.029)**	− 0.082 (0.024)**	− 0.012 (0.023)	0.000
Reimers decomposition	0.042 (0.043)	− 0.045 (0.051)			0.302
Cotton decomposition	0.040 (0.038)		− 0.042 (0.024)	0.024 (0.021)	0.047
Brown decomposition		− 0.034 (0.046)	− 0.026 (0.036)	0.110 (0.053)*	0.121
Dummy variable	0.047 (0.020)*	0.024 (0.035)	− 0.144 (0.054)**	− 0.106 (0.036)**	0.000
IV	− 0.030 (0.037)	0.021 (0.021)			0.367

(*continued*)

Table 7. *Continued.*

	USA	Europe	Other OECD	ROW	Wald test (p)
Panel	0.038 (0.037)	− 0.099 (0.043)*		− 0.060 (0.052)	0.037
Heckman selection	− 0.006 (0.023)	0.018 (0.016)	− 0.063 (0.040)	0.023 (0.028)	0.109
No hourly wages	0.044 (0.030)	− 0.028 (0.024)	0.104 (0.024)**	0.076 (0.051)	0.000
Hourly wages constructed	− 0.022 (0.025)	0.044 (0.026)	− 0.038 (0.052)	0.023 (0.061)	0.168
Gross wages	− 0.181 (0.071)*	0.086 (0.034)*	0.026 (0.124)	− 0.209 (0.184)	0.000
Potential experience	0.012 (0.022)	0.029 (0.023)	0.066 (0.018)**	0.008 (0.042)	0.075

Variables missing in wage regression					
Experience	− 0.008 (0.035)			0.233 (0.048)**	0.000
Race or immigrant	− 0.023 (0.021)	− 0.160 (0.039)**	− 0.073 (0.047)	0.189 (0.056)**	0.000
Marital status	− 0.015 (0.020)	0.002 (0.021)	− 0.045 (0.029)	− 0.041 (0.048)	0.461
Kids	− 0.035 (0.020)	0.010 (0.025)	0.075 (0.040)	0.200 (0.076)**	0.001
Marital/kids interaction			− 0.231 (0.102)*	− 0.043 (0.105)	0.004
Training	− 0.016 (0.018)	0.124 (0.073)	− 0.163 (0.073)*	− 0.075 (0.135)	0.016
Tenure	0.011 (0.019)	0.098 (0.019)**	0.103 (0.057)	0.107 (0.040)**	0.001
Occupation	− 0.010 (0.019)	0.016 (0.023)	0.042 (0.022)	0.052 (0.042)	0.116
Industry	0.009 (0.019)	0.004 (0.018)	− 0.026 (0.027)	0.027 (0.043)	0.518
Government work	0.018 (0.027)	− 0.091 (0.024)**	0.003 (0.037)	0.024 (0.056)	0.003
Union status	0.039 (0.022)	− 0.033 (0.031)	− 0.066 (0.048)	− 0.055 (0.094)	0.024
Share of females in occupation	0.050 (0.013)**	0.006 (0.051)	0.068 (0.037)	0.000 (0.000)	0.579
Full time/Part time	0.006 (0.021)	− 0.060 (0.027)*	0.070 (0.055)	− 0.022 (0.035)	0.048

Urban	0.015	0.037	0.082	− 0.255	0.000
	(0.024)	(0.029)	(0.032)*	(0.044)**	
Region	0.023	− 0.029	− 0.069	0.161	0.000
	(0.017)	(0.016)	(0.029)*	(0.040)**	
Constant	1.218	− 0.320	0.230	0.758	
	(0.205)**	(0.193)	(0.211)	(0.280)**	
Country dummies	Yes	Yes	Yes	Yes	
Year dummies	Yes	Yes	Yes	Yes	
Observations	614	327	198	346	
Adjusted R^2	0.56	0.55	0.70	0.75	

Note: Robust standard errors are expressed in parentheses, estimates are weighted with number of estimates in the paper (as in column (2) in Table 3). Wald test for equality of coefficients across regions (p). * and ** indicate significance at a 5% and 1% level, respectively

data, the unexplained gender wage gap declines. If the wage data are not per hour, but per year, month, or week this increases the gender wage gap in the non-US, non-European OECD countries, as expected. Missing control for race or immigration status has different effects for Europe and the developing world, which is not surprising, considering the entirely different groups of people that are referred to by this variable. The use of potential instead of actual experience increases the gender wage gap for the non-US, non-European OECD countries as would be expected.

Neglecting the tenure variable in the regression analysis increases the gender wage gap all over the world, even if not always significantly so. Not including the share of females in an occupation has the clearest positive effect on the gender wage gap in the US.

Female authorship, in fact, increases the estimates in Europe and the developing countries, in stark contrast to Stanley and Jarrell's (1998, 2004) results of the US. However, we corroborate their findings that female authors report significantly lower wage gaps in the US.

7. Conclusions

In this paper, we review the existing worldwide literature on the decomposition of gender wage gaps. We investigated more than 260 published papers covering 63 countries during the time period 1960s – 1990s. Meta-regression analysis allows us to review and compare this vast amount of literature in a concise and systematic way. Particular emphasis is placed on a proper consideration of the quality and reliability of the underlying study which is done by a weighting with quality indicators as well as by a direct inclusion of quality indicators in the meta-regression analysis.

There is much discussion about how to ideally investigate discrimination in wages. Meta-analysis cannot answer this question, but provides an estimate of how certain restrictions in a particular data set or the choice of a particular specification will affect the results. Our results show that data restrictions have the biggest impact on the resulting gender wage gap. Generally, studies using data sets which are limited to particular subgroups (to never-married workers, new

entrants in the labor market, or workers in narrow occupations) and therefore provide the researcher with a better comparability of the productivity of workers end up with lower gender wage gaps. In contrast to these strong results, the choice of econometric methods is less important as it concerns the concrete decomposition technique or the use of more advanced methods in the wage regressions. Frequently, researchers do not have hourly wages or actual experience at their disposal, let alone a complete record of human capital characteristics, like training on-the-job or job tenure with the actual employer. Missing or imprecise data on these human capital factors can result in serious biases in the calculation of the discrimination component. For example, in the fixed effects regressions we find that studies where work experience is missing seriously overestimate the unexplained gender wage gap. A similar problem arises, if no hourly wages are available and they have to be substituted with monthly or annual earnings, which are contaminated with labor market interruptions. Our study also found no big differences of how certain meta-independent variables affect the calculated gender wage gap in different regions of the world.

Furthermore, our analysis allowed us to investigate the gender wage gap over time. From the 1960s to the 1990s, raw wage differentials worldwide have fallen substantially from around 65 to only 30%. The bulk of this decline, however, must be attributed to better labor market endowments of females which came about by better education, training, and work attachment. Looking at the published estimates for the discrimination (or unexplained) component of the wage gap yields a less promising perspective: There is no decline over time. However, these published estimates are based on different methods and data sources. Our meta-regression analysis allows to construct a specification for a standardized gender wage gap study: applying such a unique specification – concerning data selection as well as econometric method – gives rise to a slightly more optimistic picture: The ratio of what women would earn absent of discrimination relative to their actual wages decreased approximately by 0.17% annually. This indicates that a continuous, even if moderate, equalization between the sexes is taking place.

Acknowledgements

Notes

1. See e.g. Cain (1986) and Altonji and Blank (1999) for authoritative surveys.
2. See e.g. the discussion between Hanushek (1998) and Krueger (2003).
3. Often authors also report a 'discrimination index' which is given by $D = e^U - 1$ and indicates how much higher the average female wage would be if women's endowments would be remunerated such as men's.
4. For extensions of the B-O decomposition see e.g. Brown *et al.* (1980), Reimers (1983), Cotton (1988), and Neumark (1988).
5. See e.g. Cain (1986) for a narrative overview.
6. See Weichselbaumer and Winter-Ebmer (2003b) for an analysis of the change of terminology over time.
7. E.g. O'Neill (1985) has argued that women may not be able to devote the same effort to market work as men due to household responsibilities. This may make them choose qualitatively different jobs. Light and Ureta (1995) again have found that full characterization of previous past employment experience (including the fraction of time an individual has worked during each year of the career) substantially reduces the male–female wage gap.
8. Non-English-language papers can be equally found with this strategy because in the EconLit titles are also given in English.
9. A full list of papers included in the meta-study is available from the following URL: www.econ.jku.at/weichsel/work/meta_papers.pdf.
10. Bijmolt and Pieters (2001) show in a Monte-Carlo study on meta-analyzing the effects of marketing measures that using only a single value for each study might lead to misleading results.
11. The Grubbs test (Grubbs, 1969) was used for outlier detection, which eliminates one outlier at a time. Application of this procedure led to the removal of five observations.
12. This resulted in 41 studies for the period 1959–1986 in their meta-analysis.
13. See Silber and Weber (1999) for a bootstrap approach to construct standard errors for different decomposition procedures.
14. One way to tackle the non-independence of data points is to use a different weighting scheme. We recalculated our results from Table 3 – using as weights the inverse of the number estimates available per country and year and received qualitatively very similar results. A table is available upon request from the authors.
15. Differing coefficient estimates in the case of weighted least squares are an indication for misspecification of the equation. This relative consistency of estimates across specifications is therefore a reassuring sign.
16. Stanley and Jarrell (1998 and 2004) only examine studies based on data referring to the US population at large and therefore do not report any effects for data restrictions.
17. We follow Stanley and Jarrells meta-analyses (1998, 2004) in coding the absence of variables, but include a larger set of variables.
18. Stanley and Jarrell (1998) also find a similar, but slightly larger, effect from omitting experience.
19. This also includes the impact of the gender dummy for studies not applying a Blinder–Oaxaca decomposition.
20. Plotting the reported gender wage residual against the meta wage residual for different countries (data not shown) illustrates that there are only minor research differences between countries.

21. Weichselbaumer and Winter-Ebmer (2003a) examine the effect of equal treatment laws and competition on the meta-wage residual.
22. The declining use of restricted data sets is illustrated by Weichselbaumer and Winter-Ebmer (2003b).
23. Again, we weigh all estimates of one study (one country and time period) with the inverse of the number of estimates contained in one study. This resembles the weighting scheme of column (2) in Table 3.
24. Some coefficients are missing for some regions if there was no variation in the variable in the data.

References

Abreu, M., de Groot, H. L. F. and Raymond J. G. M. Florax: A meta-analysis of beta-convergence: The legendary two percent. *Journal of Economic Surveys* (in press).

Altonji, J. G. and Blank, R. M. (1999). Race and gender in the labor market. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3. Amsterdam: North-Holland, pp. 3143–3259.

Ashenfelter, O. Harmon, C. and Oosterbeek, H. (1999) A review of estimates of the schooling/earnings relationship, with tests for publication Bias. *Labour Economics* 6(4): 453–470.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

Becker, G. S. (1991). *A Treatise on the Family*. Cambridge: Harvard University Press.

Bijmolt, T. H. A. and Pieters, R. G. M. (2001) Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters* 12(2): 157–169.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8(4): 436–455.

Brown, R. S., Moon, M. and Zoloth, B. S. (1980) Incorporating occupational attainment in studies of male–female earnings differentials. *Journal of Human Resources* 40(1): 3–28.

Cain, G. G. (1986). The economic analysis of labor market discrimination: A survey. In O. Ashenfelter and R. Layard (eds), *Handbook of Labor Economics*, Vol 1. Amsterdam: North-Holland, pp. 693–785.

Cotton, J. (1988) On the decomposition of wage differentials. *Review of Economics and Statistics* 70(2): 236–243.

Froot, K. A. (1989). Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis* 24: 333–355.

Furber, C. D. and Morgan, T. M. (1987). Lessons from overviews of cardiovascular trials. *Statistics in Medicine* 6: 295–303.

Grubbs, F. (1969) Procedures for detecting outlying observations in samples. *Technometrics* 11(1): 1–21.

Hanushek, E. A. (1998). The evidence on class size, Occasional Paper 98–1, W. Allen Wallis Institute of Political Economy, University of Rochester.

Hersch, J. and Stratton, L. S. (2000). Household specialization and the male marriage wage premium. *Industrial and Labor Relations Review* 54(1): 78–95.

Jarrell, S. B. and Stanley, T. D. (2004). Declining bias and gender wage discrimination? A meta-regression analysis. *Journal of Human Resources* 39(3): 828–838.

Laband, D. N. and Piette, M. J. (1994). The relative impacts of economics journals: 1970–1990. *Journal of Economic Literature* 32(2): 640–667.

Light, A. and Ureta, M. (1995). Early-career work experience and gender wage differentials. *Journal of Labor Economics* 13(1): 121–154.

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal* 113(1): F34–F63.

Murphy, K. M. and Topel, R. H. (1985) Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* 3(4): 370–379.

Neumark, D. (1988) Employers' discriminatory behavior and the estimation of wage discrimination. *The Journal of Human Resources* 23(3): 279–295.

Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review* 14(3): 693–709.

O'Neill, J. (1985) The trend in the male–female wage gap in the United States. *Journal of Labor Economics* 3(1): S91–S116.

Reimers, C. W. (1983). Labor market discrimination against hispanic and Black Men. Review of Economics and Statistics 65(4): 570–579.

Silber, J. and Weber, M. (1999). Labor market discrimination: Are there significant differences between the various decomposition procedures? *Applied Economics* 31: 359–365.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives* 15(3): 131–150.

Stanley, T. D. Beyond publication bias. *Journal of Economic Surveys* (in press).

Stanley, T. D. and Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys* 3(2): 161–170.

Stanley, T. D. and Jarrell, S. B. (1998). Gender wage discrimination bias? A meta-regression analysis. *Journal of Human Resources* 33(4): 947–973.

Weichselbaumer, D. and Winter-Ebmer, R. (2003a). The effects of competition and equal treatment laws on the gender wage differential, CEPR Working Paper 4015, London.

Weichselbaumer, D. and Winter-Ebmer, R. (2003b) Rhetoric of economic research: The case of gender wage differentials, CEPR Working Paper 4128, London.