Inclusion of Potentially Tainted Variables in Regression Analyses for Employment Discrimination Cases

Srijati Anandat
Kevin Gilmartin††

Multiple regression analysis is being used with increasing frequency to determine liability in employment discrimination cases and has been proposed as a method for computing relief in such cases. Courts have been inconsistent, however, in their rationales for deciding that certain explanatory variables are tainted by the employer’s discrimination and should therefore be excluded from the regression model. Courts have not sufficiently distinguished measures of employees’ current job levels, initial job levels with the employer, and job levels with previous employers and have often evaluated the regression model without reference to the specific allegations (e.g., equal pay inequity vs. disparities in rates of promotion). The authors argue that even when a variable has been shown to be tainted by discrimination, its inclusion in regression analyses may help the court determine which types of employment decisions produced the disparities and what injunctive and monetary relief is most appropriate.

INTRODUCTION ............................................. 122

I. THE UTILITY OF MULTIPLE REGRESSION ANALYSIS ............................................. 125
   A. Limitations of Regression Analysis ............................................. 125
      1. Inclusion of Tainted Variables ............................................. 126
      2. Measurement Error ..................................................... 126
      3. Omission of Crucial Variables ............................................. 126
      4. Specification Error ..................................................... 127
      5. Multicollinearity ..................................................... 127

II. EXPLANATORY VARIABLES USED IN EMPLOYMENT DISCRIMINATION CASES .......... 128

† Senior Researcher, Far West Laboratory, San Francisco; B.A. 1976, M.A. 1979, Ph.D. 1985, University of California, Berkeley.

A. Choosing Explanatory Variables in Employment Litigation ............................................ 128
   1. Human Capital vs. Establishment-Oriented Regression Models ............................................ 129
   2. Job Level Variables ............................................. 130
      a. Current Job Level ........................................... 130
      b. Initial Job Level ............................................. 132
      c. Prior Job Level ............................................. 134
      d. Job Category ............................................. 135
B. Other Variables in University Employment Cases ............................................. 136
   1. “Academic Department” as an Explanatory Variable ............................................. 136
   2. Distinguishing Qualitative Factors as Explanatory Variables ............................................. 137
   3. Performance Ratings and Disciplinary Actions as Productivity Variables ............................................. 139
C. Other Variables in Employment Discrimination ............................................. 140
   1. Longevity Factors ............................................. 140
   2. Marital Status ............................................. 141
   3. Military Experience ............................................. 142

III. TAINTED VARIABLES AND ALLOCATION OF THE BURDEN OF PERSUASION ............................................. 143
A. Courts Are Inconsistent in Their Treatment of Variables as Potentially Tainted ............................................. 143
B. The Seventh Circuit Requires the Plaintiff to Establish Taint ............................................. 144

IV. REGRESSION ANALYSES AND INCLUSION OF PRE-TITLE DISCRIMINATION ............................................. 146
A. Courts Are Divided Over Treatment of Pre-Title VII Employment Practices ............................................. 146
B. Is Bazemore Conclusive Regarding Admissibility of Pre-Title VII Conduct? ............................................. 147
C. A Proposed Solution ............................................. 149

CONCLUSION ............................................. 149

INTRODUCTION

The use of multiple regression analysis1 in employment discrimina-

---

1. Multiple regression analysis is a statistical technique for evaluating the relationship between the variable of interest (the dependent variable) and variables that are believed to influence it (the independent variables). The premise underlying multiple regression analysis is that there is a mathematical formula that can predict or explain the value of the dependent variable from the independent variables. Multiple regression analyses begin with specification of the regression model, that is, identification of the explanatory variables thought likely to influence the dependent variable and the appropriate mathematical form (e.g., linear, logarithmic, curvilinear) expressing how each
tions cases is a relatively recent development. A 1975 student note in the Harvard Law Review contained one of the first suggestions that regression models be used in employment discrimination litigation. Since that time, numerous employment discrimination cases have used regression models, particularly with respect to pay, hiring, and promotion decisions. Regression analyses are used by plaintiffs to make a prima facie case, by defendants to rebut that case, or as is most often the situation, by both parties to show the merit of their respective cases.

There are several important issues surrounding the use of regression analysis in employment discrimination litigation. One central issue concerns what variables should be included in regression analyses to account for salary differences, differential hiring rates, and promotion among employees. Potentially relevant explanatory variables include employee qualifications (e.g., years of education, previous work experience, seniority) and job content factors (e.g., job position, grade level, job performance measurements, supervisor appraisals).

The relevance of these factors to employer actions is obvious. However, such variables may be unavailable, inaccurate, or unreliable. Variables under the control of the employer are particularly suspect. Some job-related variables, such as job level (or grade level), are often excluded from regression analyses in legal proceedings because they are under the employer’s control and are therefore at risk of being “tainted” by the employer’s discrimination. According to Finkelstein, a variable is considered tainted if there is a belief that the employer can shape its definition or measurement to the disadvantage of the protected group. Interestingly, Finkelstein’s definition implies that this mere possibility—rather than evidence of discrimination—is sufficient to label a factor as “tainted.” Although legal precedents offer some guidelines regarding what variables may or may not be tainted, such precedents often conflict.

---

5. Conway & Roberts, supra note 1, at 133.
and performance appraisal ratings in a regression analysis is hotly debated because of the possibility of discriminatory influence by the employer.  

The purpose of this article is to review the various issues relating to the inclusion of potentially tainted variables in regression analyses in the context of recent employment discrimination cases. Section I presents a brief overview of the utility and the limitations of multiple regression analyses in employment discrimination litigation. Section II describes the major types of explanatory variables used in regression models and the circumstances under which courts have determined these variables to be either tainted or legitimate. In that section we argue that courts have often confused current job level, initial job level, and job level with a previous employer and that the decision of whether to include or exclude a particular explanatory variable should depend on the form of discrimination that has been alleged. Section III discusses the allocation of the burden of proof with respect to tainted variables. Section IV focuses on the related issue of the inclusion of data on employees hired prior to the actionable reach of Title VII of the Civil Rights Act of 1964.

Finally, we present some conclusions regarding the use of potentially tainted variables in multiple regression analyses in employment discrimination litigation. We argue that even when a variable is known to be tainted with discrimination, the court should require its inclusion in some analyses in order to locate the source of the disparities and fashion appropriate relief. Exclusion of an important (albeit tainted) variable renders the regression model incomplete and makes it difficult to interpret the resulting group membership coefficient. Inclusion of the tainted variable, at least as part of a set of auxiliary regression analyses, allows for investigation of the relationship among the crucial variables and the ways in which they change with respect to each other. Inclusion is essential to understanding and isolating the source of observed disparities. For example, a disparity in current salaries may be due to a disparity in initial placements, a disparity in performance appraisals, a disparity in promotion rates unconnected with performance appraisals, or a disparity in salaries within job level that is unrelated to any of the other factors. By alternately including and excluding measures of initial job level, performance appraisal, and current job level, one can discover the particular source of the salary disparities. This would have obvious relevance to the shaping of appropriate injunctive and monetary relief and to determining which class members suffered from the effects of the discrimination.

---


7. A regression analysis is not required to contain all measurable variables thought to have an effect on salary; failure to include variables will normally affect the probative value of the analysis, not its admissibility. Bazemore v. Friday, 478 U.S. 385, 400 (1986). However, auxiliary analyses
I

THE UTILITY OF MULTIPLE REGRESSION ANALYSIS

In the employment discrimination context, multiple regression analysis is most frequently used to analyze the combined effects of various independent variables on individuals’ salaries. For example, it can be used to estimate the extent to which employee salary at a certain institution depends simultaneously on seniority, years of education, and protected group membership (e.g., gender or race). Assuming a linear relationship of salary with seniority, years of education, and gender and no interaction effects among the independent variables in their influence on salary, the basic multiple regression equation would assume the following form:

\[ S = A + Bx + Cy + Dz + e \]

where:

- \( S \) = salary in dollars
- \( x \) = seniority (years with the organization)
- \( y \) = years of education
- \( z \) = gender (1 for women, 0 for men)
- \( e \) = a “random-error” term that is assumed to be zero on the average and is assumed to be uncorrelated with \( x \), \( y \), and \( z \)

A, B, C, and D are numerical constants

The numerical constants are estimated from the data by minimizing the sum of squared errors between the observed and predicted values of salaries. The coefficients (\( B \), \( C \), and \( D \)) associated with each independent variable express the strength and direction of the relationship between that variable and salary, controlling for the effects on salary of each of the other independent variables. Given the definition of \( z \), the parameter \( D \) is interpreted as the average difference in salary between men and women of identical seniority and years of education. A significantly negative value for \( D \) could be interpreted as evidence of systematic discrimination against women.

A. Limitations of Regression Analysis

The ability of multiple regression to separate the influence of sex or race on salaries from other important influences on salary makes this statistical technique useful in litigation involving allegations of employment discrimination. The courts have come to recognize this primary advantage of multiple regression analysis over simple tabular analysis.\(^8\) Courts have also come to realize certain potential problems or limitations containing tainted variables create the potential for all of the measured variables to be included in a court's analysis of a discrimination allegation.

of regression analysis. Major limitations of regression analysis that courts have recognized include the following: (1) underestimation of discrimination due to inclusion of tainted variables; (2) measurement error; (3) omission of crucial variables; (4) specification error; and (5) multicollinearity.

1. Inclusion of Tainted Variables

The effects of discrimination can be underestimated when discriminatory behavior affects one or more of the explanatory variables. For example, appraisal ratings are often alleged to be tainted by discrimination. Suppose appraisal ratings have been used as a determinant of salary and that men receive higher appraisal ratings than women, due in part to discrimination. A salary regression analysis including appraisal ratings may show a significant coefficient for that tainted variable. At the same time, however, the resulting regression coefficient for gender may be nonsignificant despite the presence of discrimination. Thus, including a tainted variable in a regression analysis results in an underestimation of the effects of discrimination.

2. Measurement Error

The effect of measurement error is usually attenuation of the numerical constants in the regression equation. Productivity factors in particular are often unquantifiable or measured imprecisely and incompletely. In university employment discrimination cases, for example, number of publications has commonly been used as a measure for quality of research. Clearly, such quantification does not adequately reflect the importance or substance of the research undertaken. Thus, even when numerous explanatory variables are included in a salary regression analysis, some aspects of "true" productivity remain unmeasured, rendering the analysis unable to explain all of the observed salary variability among individuals.

3. Omission of Crucial Variables

The courts recognize that the validity of the regression analysis hinges on the choice of proper explanatory variables. Suppose some important explanatory variable, such as prior work experience, is omitted.

---

9. For one of the most complete discussions in a court decision of the advantages and limitations of regression analysis, see Vuyanich v. Republic Nat. Bank of Dallas, 505 F. Supp. at 267-79.
from a salary regression analysis. Suppose further that women have less prior work experience than men. The resulting regression coefficient for gender may be inflated in that it reflects salary disparities that are actually attributable to the omitted variable, prior work experience. For this reason, some courts have gone so far as to reject regression analyses that have omitted important explanatory variables or failed to include all possible explanatory variables.\textsuperscript{12} The Supreme Court found this practice unacceptable in \textit{Bazemore v. Friday}.\textsuperscript{13} Accordingly, the Supreme Court held that while the omission of variables from a regression analysis may render it less probative, the omission alone is an insufficient basis to render the analysis inadmissible.\textsuperscript{14}

4. \textit{Specification Error}

Specification error occurs when the form of the regression model is erroneously specified. One example of a specification error is use of a nonlogarithmic model when the relationship between the independent and dependent variables is logarithmic. In regression analysis, consideration of all possible formats of the relationship between the explanatory variables and the dependent variable is important. The appropriate mathematical form of the relationship may be linear, loglinear, or curvilinear. Variation in the dependent variable may be explained if the interaction effect of two explanatory variables is computed.

Perhaps because of the statistical background required, courts rarely discuss specification or functional form arguments.\textsuperscript{15} One exception is \textit{Presseisen v. Swarthmore College}.\textsuperscript{16} In this case, the court accepted as valid the defendants' criticism of the plaintiffs' omission of an interaction term in a salary regression analysis. The defendants, contending that the effects of academic division on salary interacted with other explanatory variables, presented their own regression analysis that included the previously omitted interaction terms. The defendants' analysis concluded there was no correlation between salary and gender.

5. \textit{Multicollinearity}

Multicollinearity, also called collinearity, refers to the problem of instability or uninterpretability of regression results when intercorrelations among explanatory variables are high. High correlations among explanatory variables make it difficult to tease out the separate effects of

\begin{itemize}
\item \textsuperscript{13} 478 U.S. 385 (1986).
\item \textsuperscript{14} Id. at 400.
\item \textsuperscript{15} Note, \textit{Title VII, Multiple Linear Regression Models, and the Courts: An Analysis}, 46 LAW \\& CONTEMP. PROBS. 283, 284-86 (1983).
\item \textsuperscript{16} 442 F. Supp. 593, 615-19 (E.D. Pa. 1977), aff'd mem., 582 F.2d 1275 (3d Cir. 1978).
\end{itemize}
each explanatory variable on the dependent variable. In *Sobel v. Yeshiva University*, for example, the court determined that the plaintiffs' attempt to control for experience in their salary regression by simultaneously including different forms of job tenure and age variables rendered the results unreliable because of collinearity. In the employment discrimination context, collinearity between protected group membership and other independent variables is a major problem. Collinearity that does not involve group membership is less of a problem because it does not reduce the precision of the group membership coefficient, which is the focus of the analysis.

In summary, regression analysis is a powerful statistical tool for use in employment discrimination litigation. However, technical understanding of the methodology and its limitations is necessary in order to correctly interpret regression results. As will be discussed in the next section, courts have credited and discredited regression results according to their different understandings of regression analysis and their various positions on the issue of tainted (or even potentially tainted) variables.

II
EXPLANATORY VARIABLES USED IN EMPLOYMENT DISCRIMINATION CASES

A. Choosing Explanatory Variables in Employment Litigation

The choice of explanatory variables in specifying the regression model is a hotly debated issue in employment discrimination cases. For example, plaintiffs and defendants generally disagree over whether it is proper to include a job level variable in regression analyses to explain differences in salary and other employment actions. Plaintiffs generally resist its inclusion on the grounds that job level is a tainted variable and its inclusion may mask any discrimination that exists. The defendants, on the other hand, usually argue that an employee's earnings obviously depend in part on his or her job level and insist on its inclusion as a proper and important explanatory variable.

Whether or not job level should be included in a regression model is

---

20. For example, a woman's salary may be comparable to those of the men at her job level. Assume, however, that this woman's qualifications warranted placing her in a higher job level at the time she was hired. When the regression analysis compares her with others in her job level, it appears that discrimination has not occurred, when in fact it has.
a crucial issue. In his review of seventeen multiple regression analyses in
class salary discrimination cases, Fogel demonstrated just how crucial
the resolution of the job level issue is to case outcome.\textsuperscript{21} Specifically, he
observed that of the eight decisions where courts found pay discrimina-
tion, only one decision credited the use of a job level variable. Con-
versely, all nine court decisions that found no pay discrimination
credited or required a job level variable. Thus, to a great extent, resolu-
tion of the question of whether to include job level is predictive of the
outcome of the case. The plaintiffs typically prevail when regression anal-
yses omitting job level are credited, and defendants typically prevail
when regressions including job level are credited. Thus, the explanatory
variables included in the regression model accepted by the court can be
the single most important issue in an employment discrimination case.

1. \textit{Human Capital vs. Establishment-Oriented Regression Models}

Plaintiffs typically base their choice of explanatory variables on a
"human capital model." The theory underlying this labor economic
model is that certain investments (i.e., human capital) an employee
brings to the job, such as education and experience, are largely determi-
native of that employee's productivity and value.\textsuperscript{22}

In contrast, defendants in employment discrimination cases typi-
cally use a more establishment-oriented pay model in specifying their re-
gression analyses. Establishment-oriented models place more emphasis
on the job to be performed than on the qualifications of individual em-
ployees.\textsuperscript{23} The resulting salary regression models are based on the estab-
lishment's procedures for salary setting. Explanatory variables under
this model generally include job tenure and job performance on the
grounds that employers increase pay as a function of these variables. The
variable most representative of the establishment-oriented approach,
however, is job level. Job level refers to the grade level or rank of an
employee within an establishment's work force hierarchy. In a given es-
ablishment, pay rates are assigned to jobs, and jobs with similar pay
rates are typically grouped into a job level or pay grade with an estab-
lished standard base pay rate and range.

Human capital variables and establishment-oriented variables are
not incompatible in regression analyses. Many regression models
presented in Title VII cases have contained both establishment and
human capital variables. However, a major distinction between the two
models relates to the treatment of the job level variable.\textsuperscript{24} This variable

\begin{itemize}
  \item \textsuperscript{21} Fogel, \textit{supra} note 3, at 314-17.
  \item \textsuperscript{22} \textit{Id.} at 304.
  \item \textsuperscript{23} \textit{Id.} at 305.
  \item \textsuperscript{24} \textit{Id.} at 313-21.
\end{itemize}
is generally excluded in a human capital model and included in an establishment-oriented model.

Most conflicts over the treatment of a potentially tainted variable concern some type of job level variable. Accordingly, the description of the major types of explanatory variables considered in regression analyses will begin with the commonly used forms of job level variables.

2. Job Level Variables

The job level variable\(^{25}\) has assumed different forms in employment discrimination cases. The major forms include current job or grade level, initial job level with the present employer, and previous job level before being hired by the present employer. Job category, which is related to but distinct from job level, has also been used in some multiple regression analyses.

In considering whether or not to allow the inclusion of job level in salary regressions, courts need to consider regression analyses relative to the specific allegations they were designed to address. Some courts appear not to distinguish between discrimination in salary, promotions, and job allocation in crediting and discrediting certain regression analyses. If the allegation is of salary disparities between similarly situated employees, current job level is an important explanatory variable. However, if the issue is discrimination in promotions or job allocation and salary is being used as a proxy for career progression, then it may not make sense to include current job level as an independent variable. Nevertheless, inclusion of initial job level or prior job level may be appropriate even when inclusion of current job level is not. Court decisions with respect to each type of job variable are discussed in turn.

a. Current Job Level

The most commonly used form of the job level variable in regression analyses is the current job level of an employee. In *Valentino v. United States Postal Service*,\(^{26}\) where the Postal Service faced a charge of sex discrimination, the court gave more weight to the defendants' salary regressions, which included current job level, than to the plaintiffs' analyses which omitted this and other variables. The plaintiffs asserted that job level was tainted but offered no evidence to back up this assertion. The court credited inclusion of job level in the absence of tangible evidence of discrimination. The court was impressed by the ability of the defendants' regression analysis to account for much more of the variance in salary than the plaintiffs' analysis. The reported \(R^2\) was .88 for the defendants' regression as compared with .28 for the plaintiffs' regression analysis.

---

25. The terms job level, grade level, and rank are used interchangeably in this discussion.
analysis, indicating that the defendants’ analysis accounted for almost ninety percent of the variance in salaries while the plaintiffs’ analysis accounted for only about thirty percent.\(^{27}\)

The district court in *Sobel v. Yeshiva University*\(^{28}\) also found the plaintiffs’ regression analyses, which omitted rank as an explanatory variable, insufficient to support a finding of employment discrimination. *Sobel* involved allegations brought by female doctors against the university medical school. The plaintiffs argued that because academic rank is prone to being tainted, this factor should be included as an explanatory variable only where there is evidence that neutral and objective standards have been strictly followed and evidence that the decisions regarding rank were not influenced by sex discrimination. The defendants’ evidence satisfied the court, however, that the defendants handled promotions in rank in a nondiscriminatory manner. Thus, the court agreed that by omitting rank from their analyses the plaintiffs had failed to account adequately for productivity. This, the court said, resulted in an overestimation of the sex coefficient and a distortion of the regression results. In requiring the inclusion of rank, the court expressed its unusual view of rank as an appropriate proxy for productivity and quality in research and teaching. It held that “rank appears to be the only available surrogate for productivity variables that cannot be quantified in any other manner.”\(^{29}\)

In *Presseisen v. Swarthmore College*,\(^ {30}\) the court partially justified its rejection of the plaintiffs’ regression analyses, which excluded rank, by reference to the defendants’ showing that there was no statistically significant difference between men and women in the length of time for promotion from rank to rank or in the length of time within rank at Swarthmore. In addition to finding no discrimination in promotions and salary, the court also found that the university had no “policy of discriminating against women with respect to hiring; that is, with respect to hiring the overall faculty.”\(^ {31}\) Still, the court admitted to being “troubled by the fact that no women had ever been hired initially as an Associate Professor or full Professor.”\(^ {32}\) Thus, while the court alluded to the possibility of discrimination in assignment of initial rank, it did not view this unproven possibility as justification for exclusion of rank from regression analysis. The court might have been better served by a more informative series of analyses that alternately included and excluded measures of previous rank (or level of schooling), initial rank, and current rank in order

\(^{27}\) 511 F. Supp. at 944.
\(^{28}\) 566 F. Supp. at 1180-81.
\(^{29}\) *Id.* at 1180.
\(^{30}\) 442 F. Supp. at 614-17, 619.
\(^{31}\) *Id.* at 621.
\(^{32}\) *Id.*
to determine which employment decisions, if any, were producing disparities (i.e., initial placement decisions, promotion decisions, or pay decisions within rank).

Similarly, the court in *Ottaviani v. University of New York* 33 considered the possibility of taint in initial rank prior to ruling on the use of rank in a salary regression. This university sex discrimination case involved a number of charges, including discrimination in initial placement in rank. Upon finding that the defendants' statistical evidence proved there was no discrimination either in placement into initial rank or in promotion, the district court ruled that rank was a proper variable to include in regression analyses. 34 The court found plaintiffs' analyses, which excluded rank, insufficient to demonstrate sex discrimination. 35

When faced with evidence of discrimination in hiring or promotion, courts often discredit analyses that include job level. For example, in *Craik v. Minnesota State University Board*, 36 the plaintiffs presented multiple regression analyses of rank including explanatory factors such as highest academic degree, time elapsed since highest degree, and experience. The results indicated that women held lower academic rank at a statistically significant rate. 37 Consequently, the court found defendants' salary analyses including rank to be deficient because plaintiffs had already proven sex discrimination with respect to rank. However, because it required exclusion of rank from the regression analyses, the court was unable to differentiate discrimination in promotions from discrimination in salaries within rank.

b. *Initial Job Level*

Initial job level, that is, an employee's job level when hired by the present employer, is sometimes included as an explanatory variable. However, as with current job level, courts have not been consistent in their treatment of initial job level as an explanatory variable.

It has been suggested that initial job level may be preferable to current job level as an explanatory variable because the latter is potentially tainted. 38 This argument misses the mark; either current job level or initial job level may be tainted, since both are under the control of the current employer. However, a regression analysis that includes current job

34. 679 F. Supp. at 306.
35. Id. at 308.
36. 731 F.2d 465, 475-78 (8th Cir. 1984).
37. The defendants justifiably criticized the plaintiffs' use of regression as technically inappropriate where plaintiffs represented rank by artificial values of one to four as the dependent variable. However, the defendants' criticisms failed to persuade the court. Id. at 476-77.
38. Smith & Abram, supra note 11, at 68 n.144.
level tests for pay inequities within job level, while a regression analysis
tests for discrimination in promotions and other employment decisions that affect salary progression after hiring.
The court investigates both forms of discrimination in many if not most Title VII cases. Analyses should therefore be performed both with and without initial job level and current job level as explanatory variables.

In *Pouncy v. Prudential Insurance Co.*, 39 which involved allegations of racial discrimination, the court criticized the plaintiffs' salary analyses40 and promotion analyses41 because they failed to control for level of hire. The court stated that in failing to consider the level at which each employee was hired, the plaintiffs' exhibit relied upon the assumption that the average salary of all races should be the same, regardless of level of hire.42 The court preferred the defendants' salary regression analyses, which included current job level and tenure as explanatory variables.43 The court apparently viewed job level and tenure as appropriate substitutes for level of hire. Interestingly, the court noted the possibility of bias or taint in explanatory factors such as education, skill level, previous training, and supervisory appraisals, commenting on their exclusion from the defendants' analyses.44 The court did not, however, appear to consider the possibility of taint with respect to initial job level. A series of analyses designed to isolate disparities in initial placements (given prehire qualifications), disparities in promotions, and disparities in the salaries of similarly situated black and white employees could have served as the basis for a better informed opinion.

In contrast, the court in *Greenspan v. Automobile Club of Michigan*45 seemed to begin with the assumption that company-related variables, such as initial job level, are potentially tainted. The court agreed that plaintiffs appropriately omitted such variables from their regression analyses. However, unlike many cases where courts accepted one party's regression analyses and rejected the other party's analyses, this court also credited the regressions presented by the defendants, which included company-related variables. The court reasoned that there was no inconsistency in accepting both types of analyses as evidence, since they addressed different questions.46 This point of view is commendable. However, the various questions to be addressed should be determined prior to trial, and both sides should be required to design analyses to

---

40. 499 F. Supp. at 449.
41. Id. at 453.
42. Id. at 449.
43. Id.
44. Id. at 450 n.14.
46. Id. at 1064-65.
address each of the questions (e.g., discrimination in initial placements, promotions, and salary).

Unlike either the Pouncy court or the Greenspan court, the court in Chang v. University of Rhode Island\(^4\) considered evidence of possible taint prior to deciding whether or not to allow the inclusion of initial grade level. This court credited the plaintiffs' regressions of salary at hire, which excluded faculty rank, after being persuaded by plaintiffs' analyses showing disparities in initial rank placement.\(^4\) As in Craik, the court would have been better informed had it been able to compare otherwise identical analyses that alternately included and excluded measures of rank and other potentially tainted variables.

c. Prior Job Level

Previous job level, the position attained at another organization, has been used in regression analyses related to allegations of sex discrimination in university settings. Since previous job level is not under the control of the current employer, it cannot be tainted by the current employer's discrimination (although it may reflect the effects of discrimination by other employers or by society in general). In Coser v. Moore,\(^4\) the district court found probative the defendants' regression analyses of initial rank at hire, salary, and award of tenure. These included, among other explanatory variables, rank, tenure, and position at a previous institution.\(^5\) In finding the plaintiffs' regression analyses inadequate, the court noted:

> Of particular significance here, and explanatory of much of the difference between plaintiffs' and defendants' experts is plaintiffs' omission of any variable to account for prior experience. Plaintiffs sought to justify that omission on the ground that to recognize it would merely perpetuate the effects of sex discrimination that originally occurred elsewhere. However, as defendants correctly point out, in this action it is only the defendants' neutrality which is at issue, not the world's.\(^5\)

Thus, the court viewed previous rank as a legitimate explanatory variable despite the possibility of discrimination in a previous organization. The appellate court subsequently affirmed the trial court's finding for the defendants and ruled that the trial court had properly determined that prior rank and experience were appropriate explanatory variables.\(^5\)

In Ottaviani v. University of New York,\(^5\) the court found initial

---

47. 606 F. Supp. 1161, 1210-16 (D.R.I. 1985).
48. Id. at 1215-16.
50. The court's only stated caution about use of prior rank was that it would not approve the use of prior rank as the "sole determining factor in initial placement." 587 F. Supp. at 589.
51. Id. at 599.
52. 739 F.2d 746, 752 (2d Cir. 1984).
53. 679 F. Supp. at 306.
rank, current rank, and prior rank important factors in accounting for faculty employment decisions. The court, therefore, found that plaintiffs' regression analyses, which excluded these factors, were insufficient to demonstrate sex discrimination with respect to salary, rank, and initial placement in rank.\textsuperscript{54}

Although prior job level has been used in regression analyses, the use of prior \textit{salary} has yet to be fully tested in a judicial setting.\textsuperscript{55} Obviously, previous salary can be an important consideration in the setting of an employee's initial salary, especially one with some years of previous experience. How this variable would hold up to charges of bias,\textsuperscript{56} however, is yet to be seen.

d. \textit{Job Category}

Job category refers to the type of occupation, such as professor, construction worker, or secretary, rather than to the rank or grade level within the employer's work force. It has been argued that job category is less prone to taint than job level because it is determined largely by forces external to an employer.\textsuperscript{57} For example, a professor and a maintenance worker at a university have differing education and training, which determine their differing job categories. In contrast, job level is more under the employer's control, hence more likely to be tainted.

In a number of cases, plaintiffs have excluded job-category variables from their regression analyses. They often give reasons other than the tainted variable argument for such exclusions.\textsuperscript{58} In \textit{Eastland v. Tennessee Valley Authority},\textsuperscript{59} for example, the plaintiffs argued that a breakdown by job category would fragment the data base, making it more difficult to establish statistical significance. More often, the plaintiffs contend that the job categories under consideration are the same or require similar qualifications and therefore need not be analyzed separately.\textsuperscript{60}

The inclusion of job category in regression analysis has not been as

\textsuperscript{54} Id. at 308.
\textsuperscript{56} Prior salary cannot be \textit{tainted}, if "tainted" is used properly to refer to employer-controlled variables that have been affected by the employer's discriminatory behavior. Prior salary can be \textit{biased}, however, if the same amount of salary corresponds to different levels of prior experience for class members and non-class members.
\textsuperscript{57} Note, \textit{supra} note 15, at 288-92.
highly charged an issue as the inclusion of job level. This is not to say, however, that job category has never been found to be tainted. In *James v. Stockham Valves and Fittings Co.*, the appellate court criticized the defendants' salary regressions for including skill level among other variables. The problem with skill level, according to the court, was that it was a job category variable. The plaintiffs had demonstrated that the defendants had systematically excluded blacks from training and promotion for certain job categories.

In summary, courts generally credit inclusion of job level in pay equity cases. A pay equity case is a special situation wherein the plaintiffs allege discrimination in pay for similar or identical jobs but do not allege discrimination in initial placement or promotion. Thus, analyses would normally be performed within job levels or, alternatively, job level would be included as an independent variable in the regression model. For cases involving allegations of discrimination in placement or promotion as well as salary, however, use of rank is often more difficult to resolve. Most courts decide this issue on the basis of whether or not discrimination in job level and placement has been demonstrated by other evidence. Absent a showing that discrimination exists in promotion or ranking practices, many courts appear to favor inclusion of job level as an explanatory variable. However, when evidence of discrimination in job level or placement is demonstrated, courts tend to discredit regression analyses including job level. A more informative approach would be to perform the analyses both ways and compare the results to determine the effects of initial placement and promotion decisions.

**B. Other Variables in University Employment Cases**

Explanatory variables other than job level have also been the subject of the tainted variable controversy in university employment discrimination cases. These variables include academic department, quality of teaching, and quality of research.

1. “*Academic Department*” as an Explanatory Variable

Estimating the effect of academic department on salary has been the most frequent means of accounting for market factors in attempts to explain variance in faculty salaries. Due partially to outside market fac-

---

61. 559 F.2d at 332-33.
62. *Id.* at 332.
63. *Id.* at 332-33.
64. *See* *e.g.*, Valentino *v.* U.S. Postal Serv., 511 F. Supp. at 944-45.
66. *Id.* at 316.
67. *Id.* at 315.
tors, it typically takes more money to hire a professor of statistics or computer science than a professor of French or history. Thus, if women tend to be concentrated in fields characterized by lower market demand, certain observed salary differences may be attributable to such differential market conditions.

Although the influence of outside market conditions on salaries is readily apparent, the court in *Melani v. Board of Higher Education of New York*\(^69\) denied the defendants' challenge to the plaintiffs' regression analyses because they omitted academic department. In so doing, Judge Gagliardi cited the defendants' expert testimony that inclusion of academic department as an explanatory variable would not be useful due to the large number of departments.\(^70\) The judge also denied the challenge on the grounds that it is not necessary to account for every factor in a salary analysis.\(^71\)

Other courts have found inclusion of academic department in a salary regression analysis to be crucial. The court in *Wilkins v. University of Houston*\(^72\) found the plaintiffs' salary regression analysis flawed primarily because it omitted academic department. The court said that the department in which a professor teaches is the most important determining factor of salary. Likewise, the court in *Coser v. Moore*\(^73\) considered the overly broad aggregation of academic divisions in plaintiffs' salary analysis to be a major methodological deficiency. Finally, the court in *Sobel v. Yeshiva University*\(^74\) suggested that the appropriate analysis in a medical school setting would include not only department but specialties within departments as well.

2. **Distinguishing Qualitative Factors as Explanatory Variables**

While qualitative factors such as teaching quality and quality of research obviously affect faculty salaries, various problems surround their measurement. For example, quality of teaching can be measured by department chairpersons' ratings or by students' ratings. Plaintiffs may object to chairpersons' ratings because chairpersons tend to be white males and may even be defendants. Thus, these ratings may be tainted. Student ratings, on the other hand, may be considered inaccurate to the extent that they are reflective of factors such as the teacher's personality rather than teaching quality.

As an explanatory variable, quality of research faces similar quantification problems. The number of publications, the most common mea-

\(^{70}\) 561 F. Supp. at 779.
\(^{71}\) Id.
\(^{72}\) 654 F.2d 388, 402 (5th Cir. 1981), reh'g denied, 459 U.S. 822 (1982).
\(^{73}\) 587 F. Supp. at 601.
\(^{74}\) 566 F. Supp. at 1178-79.
sure of this factor, does not reflect the substantiality or importance of the publications. Moreover, there is a practical obstacle to obtaining this information; lists of publications are typically taken from curricula vitae, many of which are not updated in a timely fashion.

Recognizing that the potential for these productivity variables might incorporate the effects of discrimination, the Melani v. Board of Higher Education of New York75 court did not require the plaintiffs to include variables measuring teaching or research quality in their regression analyses. The court explained: "[A]s a result of discrimination, a woman may be assigned greater teaching responsibilities than a similarly situated man. Her ability to do research, her teaching performance, and her time for administrative work all may suffer as a result of the initial discriminatory assignment."76

Most courts, however, have been more critical of the omission of productivity variables from regression analyses. In Mecklenburg v. Montana Board of Regents of Higher Education,77 the court noted limitations resulting from the exclusion of these variables. The Sobel v. Yeshiva University78 court took a stronger stand on the issue, contending that the omission of quality of research and teaching in the plaintiffs' regression analyses was a serious inadequacy. As was discussed previously, the Sobel court held that rank should be included as a proxy in the regression analyses to avoid the inadequacy resulting from the omission of quality of research and teaching. The court reasoned that "[A]lthough rank does not completely reflect such amorphous considerations as the quality of research and teaching, clinical expertise, and reputation, it is the only available variable that seems to give any weight to these important factors."79

Even if a productivity factor is suspected or known to be tainted by the employer's discrimination, alternately including and excluding the measure will indicate the amount of influence that factor has on salaries and may help the court to focus on the aspects of the employment situation that are causing disparities. For example, ratings by department heads of teaching quality may be considered to be a tainted variable, but adding the measure to a regression model may have little or no effect on the estimated size of a salary disparity. In such a case, attention can be shifted from the ratings to other possible sources of the disparity.

75. 561 F. Supp. at 778.
76. Id.
78. 566 F. Supp. at 1179.
79. Id. at 1180 (footnote omitted).
3. **Performance Ratings and Disciplinary Actions as Productivity Variables**

While productivity in a university setting is generally reflected by teaching and research quality, typical productivity factors that have been considered for inclusion in non-university employment discrimination cases include performance ratings, disciplinary actions, and measures of leadership, skill, and output. These variables are susceptible to the same criticisms as their university counterparts, namely, that they are difficult to measure or are particularly prone to taint. Two commonly included productivity variables, performance ratings and disciplinary actions, are discussed below.

Decisions regarding inclusion of performance ratings in regression and other analyses vary. Some courts have excluded performance ratings simply because they are subjective. In *Stastny v. Southern Bell Telephone & Telegraph Co.*[^80] the trial court discounted the defendants' appraisal system as a possible explanatory factor for promotions not only because of the subjective nature of appraisals, but also because of other related practices at Southern Bell that suggested the potential for discrimination (i.e., lack of public postings of vacancies and reliance on the grapevine for news of promotion opportunities). The court stated that "[t]he operation of an opinion-based appraisal system, largely controlled by one sex, such as the one at Southern Bell, provides an ideal environment for disparate treatment of sexes."[^82] The court then cited statistical findings that indicated males received higher appraisal ratings and more promotions than females as sufficient evidence that appraisals reflect discrimination.[^83]

The court in *EEOC v. International Business Machines Co.*,[^84] on the other hand, found IBM's subjective performance evaluation system non-discriminatory despite the fact that black employees received consistently lower ratings than white employees. The IBM system impressed the court with its built-in safeguards against discrimination such as training of supervisors in how to make unbiased ratings, reviews of performance ratings by second-line supervisors, appeal procedures, and other ways in which employees could express displeasure.[^85] Thus, the court approved regression analyses including performance appraisals as an explanatory

[^82]: 458 F. Supp. at 331.
[^83]: *Id.*
[^85]: *Id.*
variable. In this particular case, all discrimination charges were dismissed.

Charges of discrimination in employment can focus not only on salary, hiring, and promotion, but also on discharge. For example, in *Coates v. Johnson & Johnson*, black former employees brought a suit attempting to show classwide discriminatory discharge. Johnson and Johnson's rebuttal included a regression analysis showing that the disparity in discharge rates could be explained by employees' disciplinary records. The plaintiffs claimed that disciplinary actions were a reflection of the employer's discrimination, but the appellate court found disciplinary actions to be an appropriate explanatory variable because no evidence was presented showing that the defendants had discriminatorily taken disciplinary actions. Thus, the court placed the burden of demonstrating that a variable is in fact tainted on the plaintiffs.

C. Other Variables in Employment Discrimination

1. Longevity Factors

The existence of a positive relationship between salary and longevity in employment is generally accepted. Regression analyses invariably include at least one longevity factor, such as years of service at the establishment, years of experience in the particular job or profession, age, or years since training. Whereas inclusion of most job tenure variables is seldom challenged, inclusion of age is open to various types of criticism. As described previously, the court in *Sobel v. Yeshiva University* found that the plaintiffs' salary regressions, which simultaneously included different forms of job tenure and age variables, were unreliable because of collinearity. Other court cases that considered regression analyses containing both types of measures, however, did not raise the issue of collinearity between job tenure and age variables.

A more substantial objection to the inclusion of age as a proxy for experience in a regression analysis is that it is likely to be misleading; it assumes that employees of the same age in the protected and unprotected groups have the same amount of experience. In *Key v. Gillette Co.*, the court criticized the plaintiffs' use of age at time of hire as a measure of

---

86. 756 F.2d 524 (7th Cir. 1985).
87. Id. at 537-38.
88. Id. at 538-39, 542-45.
89. Id. at 544-45.
90. Bodner, supra note 55, at 311-12.
91. 566 F. Supp. at 1181.
experience, asserting that men generally possess more experience than women of the same age. The defendants in Vuyanich v. Republic National Bank of Dallas\textsuperscript{94} used descriptive statistics to show that age was an inflated measure of work experience for women at their institution. The court found this evidence sufficient to render the age proxy unreliable and call into question the results of the plaintiffs' regression results.\textsuperscript{95}

The trial court in Trout v. Hidalgo\textsuperscript{96} reached a different decision about an age-based explanatory variable. This court accepted plaintiffs' salary regressions, which included an experience factor that was defined as the entire period between completion of education and hiring by the government.\textsuperscript{97} The defendants objected that this variable reflects age more than experience. They presented census data and testimony to support their contention that technical, computer-related experience is not equally distributed between men and women in the general population and that women, on average, possess less relevant experience than men of the same age.\textsuperscript{98} The court, however, found this evidence insufficient to undermine the plaintiffs' analysis, largely because the evidence was not specific to the defendants' work force.\textsuperscript{99}

Unlike the Trout court, the court in Chang v. University of Rhode Island\textsuperscript{100} discredited the defendants' inclusion of years since degree as a proxy for experience. However, a direct measure of years of experience was available in this case, and the court ruled that the use of a "tenuous" proxy measure instead of the direct measure defied logic.\textsuperscript{101}

2. Marital Status

Marital status has been used in regression analyses, presumably as a proxy for attachment to the labor force. The usual rationale for including marital status (and sometimes a variable for children) as an explanatory variable is the belief that married men are more attached to the labor force than married women and are therefore more productive and less likely to leave the employer.\textsuperscript{102} According to this view, the lower average salary of women can be partially explained by their lower pro-

\textsuperscript{94} 505 F. Supp. at 311-13.
\textsuperscript{95} Id. at 314-17.
\textsuperscript{97} 517 F. Supp. at 878-79, 884.
\textsuperscript{98} Id. at 881.
\textsuperscript{99} Id. Interestingly, during the relief phase of the Trout case, a special master appointed by the court disapproved the use of the age proxy. 53 Empl. Prac. Dec. (CCH) ¶ 39,924 at 62,506-08 (D.D.C. 1990).
\textsuperscript{100} 606 F. Supp. at 1216.
\textsuperscript{101} Id.
\textsuperscript{102} Finkelstein, supra note 4, at 740.
ductivity. In *Stastny v. Southern Bell Telephone & Telegraph Co.*,\(^{103}\) the court found some of the explanatory variables in the employer's salary analyses, including marital status, suspect. The court noted that "there is no evidence that an unmarried [sic] woman is less likely to be a productive worker than a married man"\(^{104}\) and held that marital status is an automatically tainted variable that should not be included in regression analyses.\(^{105}\)

Finkelstein described the results of one case study of a large metropolitan publishing firm in which the investigator concluded that a major source of salary disparities was the belief of managers that married men deserved higher earnings than married women.\(^{106}\) Marital status is generally not included in multiple regression analyses. When attachment to the labor force is included in an analysis (usually in econometric or sociological studies, rather than in employment discrimination cases), it is usually measured with more direct proxies such as absentee rates and periods of part-time employment or unemployment.

3. **Military Experience**

As was previously discussed, the inclusion of a factor that is correlated with sex or race in a regression analysis can reduce or eliminate the statistical significance of group membership. In *Rossini v. Ogilvy & Mather Inc.*,\(^{107}\) the plaintiffs criticized the defendants' regression analyses for including military experience as an explanatory variable, asserting that this variable was included merely to reduce the difference between men's and women's salaries, since far more men than women have military experience.\(^{108}\) Despite the fact that there was no evidence indicating that this advertising agency actually considered military experience in job decisions, the district court found that there was a sufficient basis for including military experience as an explanatory variable.\(^{109}\) Moreover, the court went so far as to reject one of the plaintiffs' statistical tables on the speculative basis that the results might possibly have been different if military experience had been included.\(^{110}\) The appellate court later ruled that the lower court had abused its discretion by rejecting the plaintiffs'
statistical results for failing to include military experience. However, it also found that the decision to credit the defendants' analyses that included military experience was not clearly erroneous.

III
TAINTED VARIABLES AND ALLOCATION OF THE BURDEN OF PERSUASION

As indicated in the preceding section, courts tend to discredit or reject regression analyses that include as explanatory factors variables for which there is some evidence of taint. As long as this tendency continues, the allocation of the burden of proof with respect to the use of potentially tainted explanatory factors remains an important issue. Courts have been inconsistent in their allocation of the burden of proof in such cases. Some courts have treated certain variables as automatically tainted. Others have required defendants show that a variable is untainted before using it. Still others have considered it the plaintiffs' responsibility to show that a variable used by the defendants is tainted.

A. Courts Are Inconsistent in Their Treatment of Variables as Potentially Tainted

The reasoning in Melani v. Board of Higher Education of New York is based on a lack of differentiation between tainted variables (known to reflect the effects of discriminatory acts by the employer) and potentially tainted variables (merely under some amount of control by the employer, with no evidence that they have been affected by discrimination). The court ruled that potentially tainted variables, including all measures of productivity, need not be included in a regression model:

Defendant claims that plaintiffs' studies attribute to sex discrimination salary differentials actually due to factors such as publications, total years of teaching experience, quality of teaching, committee work and community service. With regard to faculty productivity while at CUNY, it has been recognized that variables pertaining to an individual's accomplishments during his employment may incorporate the effects of discriminatory decisions rather than provide an independent measure of his

111. 798 F.2d 590, 603-04 (2d Cir. 1986).
112. Id.
113. In opposition to which we have argued that complete exclusion of variables reduces the amount of information available to the court.
114. See, e.g., Melani v. Board of Higher Educ. of N.Y., 561 F. Supp. at 778 (use of post-hire publications in academic salary determinations as a variable may be tainted because the level of publication may itself be reflective of employer's discriminatory actions).
116. See, e.g., Coates v. Johnson & Johnson, 756 F.2d at 544.
productivity. . . . As defendant's expert concedes, a regression analysis seeking to determine the existence of sex-based salary discrimination should not include as independent variables productivity factors which may be affected significantly by the employer's actions. Plaintiffs consequently are not required to have included in their regression models variables measuring post-hire publications, teaching quality, and committee or community work.\textsuperscript{118}

The \textit{Trout} court cited \textit{Presseisen v. Swarthmore College} in ruling against the defendants' use of initial grade level as an explanatory variable, stating that "it is commonly accepted that it is inappropriate to include as an independent variable a factor within defendants' control unless it has been established that they did not discriminate in exercising that control."\textsuperscript{119} However, \textit{Presseisen} does not support the \textit{Trout} ruling. The \textit{Presseisen} court had found no discrimination in promotion decisions and therefore ruled that it was improper for plaintiffs' expert to \textit{exclude} rank from his regression analyses;\textsuperscript{120} the court did not rule on whether it would have been proper to \textit{include} rank in the absence of evidence concerning discrimination in promotions, nor did it indicate which party would have the burden of proving whether or not rank was tainted.

Like the \textit{Trout} court, the \textit{James v. Stockham Valves \\& Fittings Co.}\textsuperscript{121} court appears to have ruled that it is the defendants' responsibility to show that explanatory variables are untainted before using them in an analysis. However, the court never explicitly addressed the burden of proof issue. Regression analyses performed by defendants' expert were criticized and discounted on several grounds, one of which was that he had made no attempt to control or check for racial bias in his productivity factors.\textsuperscript{122}

\textbf{B. The Seventh Circuit Requires the Plaintiff to Establish Taint}

The inconsistencies and flawed or implicit reasoning in the decisions discussed seem to stem in part from the fact that very few cases have explicitly considered the question of who has the burden of proof with respect to potentially tainted variables. One exception is the Seventh Circuit case of \textit{Coates v. Johnson \\& Johnson.}\textsuperscript{123} In \textit{Coates}, the variable at issue was employees' disciplinary records. The defendants' regression analyses used disciplinary records to account for apparent black-white differences in discharge rates. The plaintiffs countered that the defend-

\textsuperscript{118} \textit{Id.} at 778 (emphasis added, citations omitted). Note the court's ruling even though professional publications are only indirectly under the control of the employer and community work is probably not under the control of the employer at all.

\textsuperscript{119} 517 F. Supp. at 886 n.47.

\textsuperscript{120} \textit{Presseisen v. Swarthmore College}, 442 F. Supp. at 614.

\textsuperscript{121} 559 F.2d 310 (5th Cir. 1977).

\textsuperscript{122} \textit{Id.} at 332.

\textsuperscript{123} 756 F.2d 524 (7th Cir. 1985).
ants had also discriminated in discipline and therefore should not be allowed to include the possibly tainted factor without first demonstrating that this factor was untainted. The court disagreed with the plaintiffs and affirmed the trial court's decision in favor of the defendants. In so doing, the appellate court concluded that the defendants had successfully rebutted the plaintiffs' prima facie case by presenting regression analyses showing that the statistical disparity in discharge rates could be explained by prior disciplinary actions. The appellate court held that the plaintiffs had the burden of persuading the court that the disciplinary system was discriminatory. Thus, in deciding this case, the Seventh Circuit held that once an employer offers statistical analyses including a potentially tainted variable, the burden falls on the plaintiffs to prove the taint of that factor:

Placing the burden on the defendant would skew the general Title VII framework for allocating burdens and would be inconsistent with the principle that the plaintiffs in a Title VII case retain the ultimate burden of persuasion on the issue of discrimination. Plaintiffs should not be able to shift the burden of persuasion by alleging that some factor in defendant's control has been used by the defendant discriminatorily. . . . By including the factor and showing that the statistics no longer indicate discrimination, the defendant responds to the particular proof used by the plaintiffs to establish their prima facie case and thus raises a genuine issue of fact as to whether the apparent disparate treatment is in fact due to discrimination.

. . . A plaintiff can meet this burden by introducing evidence sufficient for the court to infer that the explanatory factor was subject to the control of the defendant and the defendant exercised that control in an unlawful, discriminatory manner. The defendant must then either rebut this evidence or, if possible, come forward with another explanation for the alleged discriminatory result. But the plaintiffs must bear the ultimate burden of persuasion when they claim that an explanatory variable included in the defendant's statistical analysis merely reflects other discrimination by the defendant.

Judge Cudahy did not agree with the majority on this point. Although concurring in the decision reached by the majority, he asserted that the burden of showing that an explanatory factor is tainted should not always be placed on the plaintiff. Judge Cudahy argued that "the defendant should not be allowed to appeal to a factor within his control as an explanation unless he can show that it is not biased, and a complete lack of evidence either way should ordinarily work to the detriment of

124. Id. at 542-43.
125. Id. at 545.
126. Id. at 544-45.
127. Id. (citations omitted).
the defendant and not the plaintiff.”

The allocation of the burden of proof concerning the taint of explanatory variables under the control of the employer becomes a crucial issue only if the court is expected to rule absolutely that a particular variable should be included or excluded in the regression analyses. However, if instead both parties are encouraged to perform several versions of their analyses, alternately including and excluding suspect variables, then the court will have information about the extent to which those variables reduce the salary disparity. When combined with other evidence as to whether those variables reflect discrimination by the employer, this information can serve as the basis for a decision concerning not only the liability of the employer but also which personnel practices produced the discriminatory effects. For example, if the Trout court had considered analyses with and without initial grade level as an explanatory variable, it would have been in a better position to distinguish the effects of initial placement decisions from the effects of subsequent promotion decisions. The distinction might have had implications for injunctive relief, for arguing which class members had suffered from the discriminatory practices, how monetary relief should be computed for them, and which types of personnel actions should be monitored by the court.

IV
Regression Analyses and Inclusion of Pre-Title VII Discrimination

Another highly relevant issue involves the inclusion of data on employees hired prior to the determinative date of Title VII (for example, March 24, 1972, for educational institutions and public employers). The question of interest is essentially whether an employer is liable if it provides equal post-Act treatment to employees but fails to rectify pre-Act disparities. Courts have been divided as to the inclusion of data from employees hired prior to the actionable reach of Title VII.

A. Courts Are Divided Over Treatment of Pre-Title VII Employment Practices

Trout v. Hidalgo illustrates the confusion among the courts over this issue. The district court credited the plaintiffs’ analyses of salaries between 1972 and 1979, which were based on regressions from a data base that included employees hired both before and after Title VII applicability. While recognizing that discriminatory actions that occurred

128. Id. at 555. In the terminology of this article, the factor would be referred to as “tainted” rather than “biased.” See supra note 56.
130. Id. at 880.
prior to 1972 are not directly actionable, the district court contended that disparities in grade level and salary that began prior to 1972 but remained afterwards could be attributable to the employer. 131

The appellate court subsequently disagreed with the lower court on this point. According to the appellate court:

The District Court concluded that plaintiffs need not factor out time-barred discrimination because defendants may be held liable for the continuing effects of that discrimination. This theory was once in vogue, but is flatly inconsistent with the Supreme Court’s pronouncements in Hazelwood School District v. United States and United Air Lines, Inc. v. Evans. 132

The issue of inclusion of pre-Act hires, however, was not resolved definitively in this case, even though the Supreme Court granted certiorari. 133 The Supreme Court simply remanded the case to the Court of Appeals for remand back to the District Court for a factual determination of the evidentiary value of respondents’ and petitioners’ statistical evidence in light of the Court of Appeals’ conclusions of law concerning employment decisions that were not actionable in this case. 134

Although the issues remain unsettled, many courts have held that the inclusion of pre-Act hires compromises the probative value of regression analyses. In Sobel v. Yeshiva University, 135 the court determined that one of the many reasons for not finding the institution guilty of gender discrimination was that it ought not be held liable for the continued post-Act effect of any pre-Act discrimination. Likewise, in Griffin v. Board of Regents of Regency Universities, 136 which involved charges of gender discrimination, the Seventh Circuit upheld the district court’s rejection of plaintiffs’ statistics, relying in part on the failure to factor out the effects of pre-1972 employment decisions.

B. Is Bazemore Conclusive Regarding Admissibility of Pre-Title VII Conduct?

One month after Griffin, the Supreme Court rendered a decision in Bazemore v. Friday, 137 which some authors believe has resolved the pre-Act discrimination issue. 138 The Bazemore case involved an allegation that the North Carolina Agricultural Service had engaged in a pattern

131. Id. at 879-80.
134. Id.
136. 795 F.2d 1281, 1289-90 (7th Cir. 1986).
138. See, e.g., Norris, supra note 19, at 92-96; Barrett & Sansonetti, supra note 58, at 507.
and practice of racial discrimination against black employees.\textsuperscript{139} Prior to 1965, the organization had consisted of a white branch and a Negro branch with a separate salary structure in each branch. The two branches were merged prior to Title VII's enactment, but wage disparities continued—black employees were paid less than white employees at the same job level.\textsuperscript{140} At trial, the plaintiffs presented salary regression analyses including race, education, tenure, and job level as explanatory variables.\textsuperscript{141} Although these analyses showed a significant negative effect for race in certain years, the district court ruled against the plaintiffs on all charges, finding that when all relevant variables were included in the analysis, the employer had maintained a nondiscriminatory policy since Title VII became effective.\textsuperscript{142} The appellate court upheld the district court's decision and ruled that the employer was under no affirmative obligation to eliminate the difference in salaries.\textsuperscript{143}

The Supreme Court unanimously reversed the appellate court's decision on salary discrimination.\textsuperscript{144} In a pointed criticism of the Fourth Circuit's conclusion that the law did not require elimination of pre-Act wage discrimination, Justice Brennan stated:

> The error of the Court of Appeals with respect to salary disparities created prior to 1972 and perpetuated thereafter is too obvious to warrant extended discussion: that the Extension Service discriminated with respect to salaries prior to the time it was covered by Title VII does not excuse perpetuating that discrimination after the Extension Service became covered by Title VII. To hold otherwise would have the effect of exempting from liability those employers who were historically the greatest offenders of the rights of blacks. \ldots

> Each week's pay check that delivers less to a black than to a similarly situated white is a wrong actionable under Title VII, regardless of the fact that this pattern was begun prior to the effective date of Title VII.\textsuperscript{145}

Interpretations of Bazemore are varied. The trial court in Trout v. Lehman,\textsuperscript{146} for example, contended that Bazemore allows for the inclusion of data on pre-Act (i.e., previously nonactionable) employment decisions, which effectively overturned United Air Lines, Inc. v. Evans.\textsuperscript{147} Other published analyses of Bazemore conclude that post-Act pay data on employees hired before enactment of Title VII can be used to show

\textsuperscript{139} 751 F.2d at 665.
\textsuperscript{140} Id. at 666.
\textsuperscript{141} Id. at 671.
\textsuperscript{142} Id. at 672.
\textsuperscript{143} Id. at 670.
\textsuperscript{144} 478 U.S. at 386-87.
\textsuperscript{145} Id. at 395-96 (emphasis in original).
\textsuperscript{147} 431 U.S. 553 (1977).
that discrimination exists during the actionable period. Each set of paychecks issued by the Agricultural Extension Service under Title VII represented a new discriminatory act, since pay was based on de facto separate salary structures. A logical extension of this point of view is that, if the Agricultural Extension Service had instead made adjustments to salaries so that black and white employees at the same job levels had earned approximately the same amount, then the previous discrimination would not have been actionable.

Thus, there appear to be two conflicting views on the continuing effects of pre-Act discrimination. The first view is that pre-Act employment decisions are actionable if their effects continue into the post-Act period. The second view is that the continuing effects of pre-Act discrimination are not actionable and pre-Act employment decisions should therefore be excluded from regression analyses.

C. A Proposed Solution

The different viewpoints on Bazemore and the issue of the continuing effects of pre-Act discrimination indicate that the question of inclusion in analyses of data on pre-Act employment decisions is not yet settled. Nevertheless, suppose that pre-Act employment decisions should be excluded from regression analyses. How could this be accomplished? One solution adopted in Pouncy v. Prudential Insurance Co. is to exclude pre-Act hires from the analyses. In light of Bazemore, however, total exclusion of pre-Act hires from analyses may be difficult to justify. A less extreme solution outlined by the Sobel v. Yeshiva University court is to perform separate regression analyses on pre-Act and post-Act hires. However, Finkelstein has suggested a better solution: including as an explanatory variable the employees' job levels or positions just prior to the actionable period for Title VII. In this way, the effects of pre-Act decisions can be statistically controlled, and any disparities discovered by the regression analyses must have occurred during the period of legal responsibility.

CONCLUSION

Review of the cases involving tainted variable arguments makes it clear that courts have varied in their treatment of tainted or potentially tainted variables. Nevertheless, some patterns may be discerned. Courts tend to consider job level variables to be crucial and legi-
mate explanatory variables in accounting for employee compensation. In pay equity cases, for example, courts normally approve the inclusion of job level since the allegations concern differing pay within jobs or job levels. For cases involving allegations of discrimination in placement or promotion, courts tend to favor inclusion of job level only in the absence of convincing evidence that this variable is tainted by discrimination with respect to either placement or promotions. If presented with such evidence, courts tend to discredit regression analyses that include a job level variable. More care should be taken, however, to distinguish between current job level, initial job level, and job level with a prior employer.

Courts are more cautious about inherently subjective variables such as performance appraisals. Correctly or incorrectly, a number of courts have treated such variables as automatically tainted because of the degree of subjectivity involved. Courts are also becoming more critical of age and age-based proxy variables for relevant work experience. The issue with these variables is not taint, since age is not under the control of the employer. Rather, these variables are seen as potentially inaccurate and biased (i.e., related to experience differently for class members and non-class members), especially in cases involving alleged gender discrimination.

The appropriate allocation of burden of proof with respect to potentially tainted variables remains unclear. The Seventh Circuit ruled that the plaintiff bears this burden since it is the plaintiff who bears the ultimate burden of proof. The opposing view implicit in other court decisions is that the defendant must shoulder this burden. Because the defendant is typically the one to include potentially tainted variables and such variables are typically under the employer's control, the defendant is viewed as the appropriate party to establish the legitimacy of potentially tainted variables.

Recent decisions underscore the need for courts to consider regression analyses relative to the specific allegations they were designed to address. Some courts appear not to distinguish between discrimination in salary, promotions, and job allocation in crediting and discrediting certain regression analyses. If the allegation is one of salary disparities between similarly situated employees, current job level is an important explanatory variable. However, if the allegation is of discrimination in promotions or job allocation and salary is being used as a proxy for career progression, then it may not make sense to include current job level as an independent variable. Inclusion of initial job level or prior job level may nevertheless be appropriate even when inclusion of current job level is not.

Irrespective of the question of burden and the nature of the particular variable, courts tend to approve of the omission of a tainted variable
from regression analyses in the face of tangible evidence of discrimination. From a statistical as well as a legal viewpoint, however, such an omission may be inappropriate. Aickin argues that a tainted but relevant explanatory variable should be included in regression analyses. He points out that the inclusion may not, as is presumed, result in a reduction of the statistical significance level of the regression coefficient associated with group membership. When the tainted variable is omitted from the regression equation, its effects are divided among the included explanatory variables and the error term. Thus, even though the resulting coefficient for group membership may be more negative when the tainted variable is omitted, the error term will also be inflated, with the final effect that the coefficient for group membership may be less statistically significant. Including the tainted variable may make the coefficient for group membership smaller, but may increase its statistical significance.

More importantly, inclusion of the tainted explanatory variable in some analyses can be crucial to the achievement of a more accurate assessment of the locus and extent of the discrimination. Exclusion of an important albeit tainted variable renders the regression model incomplete and makes interpretation of the resulting group membership coefficient difficult. The tainted variable should be included, at least as part of a set of auxiliary regression analyses, so that the relationship among the crucial variables and the ways they change with respect to each other can be investigated. The inclusion is essential to understanding and isolating the source of observed disparities. For example, a disparity in current salaries may be due to a disparity in initial placements, a disparity in performance appraisals leading to lower salary increases, a disparity in promotion rates unconnected with performance appraisals, or a disparity in salaries within job level that is unrelated to any of the other factors. By alternately including and excluding measures of initial job level, performance appraisal, and current job level, the particular source of the salary disparities may be discovered. This would have obvious relevance in the shaping of appropriate injunctive and monetary relief and determining which class members suffered from the effects of the discrimination.

The use of tainted variables in multiple regression analyses in employment discrimination cases encompasses numerous technical and legal issues. Consistency among the courts in their treatment of such variables should increase somewhat as levels of familiarity with regression techniques increase. What remain to be resolved, however, are issues about the range of analysis information to be considered by the

---

court (i.e., only a single approved regression model vs. various auxiliary regression models that differ in important ways) and about the correct application of regression evidence to employment discrimination law.