ABSTRACT

Title of dissertation: ESSAYS ON FIRM COMPENSATION POLICY AND CONFIDENTIALITY PROTECTION AND IMPUTATION IN THE QUARTERLY WORKFORCE INDICATORS

Bryce Stephens, Doctor of Philosophy, 2006

Dissertation directed by: Professor John C. Haltiwanger
Department of Economics

This thesis is comprised of three chapters. The first chapter develops a statistical model that enriches the role that firms play in wage determination, allowing firms to influence both average wages as well as the returns to observable worker characteristics. I exploit the hierarchical nature of a unique employer-employee linked dataset maintained by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program, estimating a multilevel statistical model of earnings that accounts for firm-specific deviations in average wages as well as the returns to components of human capital – race, gender, education, and experience – while also controlling for person-level heterogeneity in earnings. Results suggest that there is significant variation in the returns to worker characteristics across firms. First, estimates of the parameters of the covariance matrix of firm-specific returns are statistically significant. Firms that tend to pay higher average wages also tend to pay higher than average returns to worker characteristics; firms that tend to reward

highly the human capital of men also highly reward the human capital of women. Second, the firm-specific returns account for roughly 7% of the variation in wages – approximately 30% of the variation in wages explained by firm-specific intercepts alone.

The second chapter (joint with John M. Abowd and Lars Vilhuber) and third chapter discuss the confidentiality protections and imputation of place-of-work in the Quarterly Workforce Indicators produced by the Census Bureau's Longitudinal Employer-Household Dynamics Program as a part of its Local Employment Dynamics Estimates partnership with 37 state Labor Market Information offices. Both chapters provide a discussion of methodology as well as assessments of the impact of the confidentiality protections and imputation of place-of-work on the Quarterly Workforce Indicators public use measures.

ESSAYS ON FIRM COMPENSATION POLICY AND CONFIDENTIALITY PROTECTION AND IMPUTATION IN THE QUARTERLY WORKFORCE INDICATORS

by

Bryce Stephens

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:
Professor John C. Haltiwanger, Chair/Advisor
Professor John M. Abowd, Advisor
Professor John Shea, Advisor
Professor John Chao
Professor John Iceland

DEDICATION

This thesis is dedicated to my parents Bruce and Cathy Stephens.

ACKNOWLEDGMENTS

I would like to thank the members of my dissertation committee and particularly Professors John C. Haltiwanger, John Shea, and John Abowd for their encouragement, support, and advice on all three chapters of this thesis. The first chapter benefitted greatly from comments received from participants in the macroeconomics brown bag seminar at the Department of Economics, University of Maryland. All three chapters benefitted from helpful comments from Simon Woodcock, Lars Vilhuber, and other staff at the Longitudinal Employer-Household Program.

My parents Bruce and Cathy Stephens and sister Jessica Stephens have been an endless source of support during my years in the graduate program of economics at the University of Maryland and during the time spent writing this thesis. Finally, a special thanks is due to my partner Boaz Green without whom I may have never had the courage to finish this dissertation.

TABLE OF CONTENTS

List of Tables vi

List of Figures x

1 Wage Dispersion, Compensation Policy, and the Role of Firms 1
1.1 Introduction 1
1.2 Firm Compensation Policy 6
1.3 Statistical Model 12
1.3.1 Multilevel Model 12
1.3.2 Model Estimation and Prediction 19
1.4 Data Description 20
1.5 Estimation and Results 25
1.5.1 Specifications 25
1.5.2 Results 32
1.5.2.1 Model Selection 32
1.5.2.2 Coefficient and Variance Estimates 34
1.5.2.3 Correlations 49
1.5.2.4 Analysis of Variance 52
1.6 Conclusion 55

2 Confidentiality Protection in the Quarterly Workforce Indicators 58
2.1 Introduction 58
2.2 Multiplicative noise model 62
2.3 Applying the fuzz factors to estimates 64
2.3.1 Distorting totals 65
2.3.2 Distorting averages of magnitude variables 65
2.3.3 Distorting differences of counts and magnitudes 66
2.4 Item suppression 67
2.5 Weighting the QWI 69
2.6 Extent of protection 71
2.7 Analytic validity 76
2.7.1 Time-series properties of distorted data 77
2.7.2 Cross-sectional unbiasedness of the distorted data 84
2.8 Concluding remarks 87

3 Imputation of Place of Work in the Quarterly Workforce Indicators: Implementation and Assessment 88
3.1 Introduction 88
3.2 A Probability Model for Employment Location 90
3.2.1 Definitions 90
3.2.2 The Probability Model 91
3.2.3 Estimation 92
3.3 Imputing Place of Work 93

3.3.1 Sketch of Imputation Method 93
3.3.2 Implementation 94
3.3.2.1 Establishment Data 94
3.3.2.2 Worker Data 95
3.3.2.3 Candidates 95
3.3.2.4 Imputation and Output Data 96
3.4 Analysis 98
3.4.1 Analytic Validity 98
3.4.2 Time Series Correlation 106
3.5 Conclusion 107

A REML Estimation of Variance Parameters 109
A.1 Transforming the Data 109
A.2 Estimation of Variance Components 110

B Derivation of Best Linear Unbiased Estimates (BLUEs) and Best Linear Unbiased Predictors (BLUPs) 112

C Details of Specifications 115
C.1 Specification 1 115
C.2 Specification 2 117
C.3 Specification 3 118
C.4 Specification 4 119

D Variance Estimates: Previous Results 120

E Definitions of Fundamental LEHD Concepts 121
E.1 Dates 121
E.2 Employer 121
E.3 Establishment 121
E.4 Employee 122
E.5 Job 122
E.6 Unemployment Insurance wage records (the QWI system universe) . . 123
E.7 Employment at a point in time 124
E.8 Employment for a full quarter 125
E.9 Point-in-time estimates of accession and separation 126
E.10 Individual concepts 127
E.11 Establishment concepts 133

F Bias Summary Statistics 142

LIST OF TABLES

1.1 Variables Used in Analysis 26

1.2 Sample Demographics and Job Characteristics 27

1.3 Sample Job Characteristics and Firm Interactions 28

1.4 Model Descriptions 30

1.5 Model Selection 33

1.6 Fixed Coefficient Estimates for Specifications 1 through 3 35

1.7 Fixed Coefficient Estimates for Specification 4: Women 37

1.8 Fixed Coefficient Estimates for Specification 4: Men and State Intercepts 39

1.9 OLS Coefficient Estimates Corresponding to Specification 4: Women 41

1.10 OLS Coefficient Estimates Corresponding to Specification 4: Men and State Intercepts 42

1.11 Variance Estimates: Firm and Person Intercepts and Residuals 43

1.12 Variance Estimates: Firm Components 44

1.13 Correlations for Specifications 1 and 2 50

1.14 Correlations for Specifications 3 and 4 51

1.15 Analysis of Variance 53

2.1 Disclosure flags in the QWI 69

2.2 Small Cells: B, Unweighted vs. Weighted 72

2.3 Small Cells: B, Undistorted vs. Distorted 73

2.4 Small Cells: B, Raw vs. Published 74

2.5 Distribution of the Error in the First Order Serial Correlation: SIC-division × County, Raw vs. Distorted Data 79

2.6 Distribution of the Error in the First Order Serial Correlation: County x SIC-division × County, Raw vs. Published Data 80

2.7 Distribution of the Error in the First Order Serial Correlation: Two-digit SIC × County, Raw vs. Distorted Data 81

2.8 Distribution of the Error in the First Order Serial Correlation: Two-digit SIC × County, Raw vs. Published Data 82

2.9 Distribution of Correlation Coefficients: SIC Division × County, Raw vs. Published Data 83

3.1 QWI Public Use Variables 102

3.2 Distribution of Correlation Coefficients: SIC Division × County, Unweighted MNTRUE vs. Unweighted MNIMPUTED 107

D.1 Variance Estimates: Intercepts 120

F.1 Distribution of Bias Measures for County by SIC Division: M and B . 143

F.2 Distribution of Bias Measures for County by SIC Division: E and F . 145

F.3 Distribution of Bias Measures for County by SIC Division: W1 and Z_NA 146

F.4 Distribution of Bias Measures for County by SIC Division: Z_NH and Z_NR 147

F.5 Distribution of Bias Measures for County by SIC Division: Z_NS and Z_W2 148

F.6 Distribution of Bias Measures for County by SIC Division: Z_W3 and Z_WFA 149

F.7 Distribution of Bias Measures for County by SIC Division: Z_WFS and Z_WH3 150

F.8 Distribution of Bias Measures for County by SIC Division: A and FA 151

F.9 Distribution of Bias Measures for County by SIC Division: FJC and FJD 152

F.10 Distribution of Bias Measures for County by SIC Division: FJF and FS 153

F.11 Distribution of Bias Measures for County by SIC Division: FT and H 154

F.12 Distribution of Bias Measures for County by SIC Division: H3 and JC 155

F.13 Distribution of Bias Measures for County by SIC Division: JD and JF 156

F.14 Distribution of Bias Measures for County by SIC Division: R and S . 157

F.15 Distribution of Bias Measures for County by SIC Division: Z_dWA and Z_dWS . . . 158

F.16 Distribution of Bias Measures for County by 2-Digit SIC: M and B . 160

F.17 Distribution of Bias Measures for County by 2-Digit SIC: E and F . . 161

F.18 Distribution of Bias Measures for County by 2-Digit SIC: W1 and Z_NA . . . 162

F.19 Distribution of Bias Measures for County by 2-Digit SIC: Z_NH and Z_NR . . . 163

F.20 Distribution of Bias Measures for County by 2-Digit SIC: Z_NS and Z_W2 . . . 164

F.21 Distribution of Bias Measures for County by 2-Digit SIC: Z_W3 and Z_WFA . . . 165

F.22 Distribution of Bias Measures for County by 2-Digit SIC: Z_WFS and Z_WH3 . . . 166

F.23 Distribution of Bias Measures for County by 2-Digit SIC: A and FA . 167

F.24 Distribution of Bias Measures for County by 2-Digit SIC: FJC and FJD . . . 168

F.25 Distribution of Bias Measures for County by 2-Digit SIC: FJF and FS 169

F.26 Distribution of Bias Measures for County by 2-Digit SIC: FT and H . 170

F.27 Distribution of Bias Measures for County by 2-Digit SIC: H3 and JC 171

F.28 Distribution of Bias Measures for County by 2-Digit SIC: JD and JF 172

F.29 Distribution of Bias Measures for County by 2-Digit SIC: R and S . . 173

F.30 Distribution of Bias Measures for County by 2-Digit SIC: Z_dWA and Z_dWS . . . 174

F.31 Distribution of Bias Measures for County by 3-Digit SIC: M and B . 175

F.32 Distribution of Bias Measures for County by 3-Digit SIC: E and F . . 176

F.33 Distribution of Bias Measures for County by 3-Digit SIC: W1 and Z_NA . . . 177

F.34 Distribution of Bias Measures for County by 3-Digit SIC: Z_NH and Z_NR . . . 178

F.35 Distribution of Bias Measures for County by 3-Digit SIC: Z_NS and Z_W2 . . . 179

F.36 Distribution of Bias Measures for County by 3-Digit SIC: Z_W3 and Z_WFA . . . 180

F.37 Distribution of Bias Measures for County by 3-Digit SIC: Z_WFS and Z_WH3 . . . 181

F.38 Distribution of Bias Measures for County by 3-Digit SIC: A and FA . 182

F.39 Distribution of Bias Measures for County by 3-Digit SIC: FJC and FJD . . . 183

F.40 Distribution of Bias Measures for County by 3-Digit SIC: FJF and FS 184

F.41 Distribution of Bias Measures for County by 3-Digit SIC: FT and H . 185

F.42 Distribution of Bias Measures for County by 3-Digit SIC: H3 and JC 186

F.43 Distribution of Bias Measures for County by 3-Digit SIC: JD and JF 187

F.44 Distribution of Bias Measures for County by 3-Digit SIC: R and S . . 188

F.45 Distribution of Bias Measures for County by 3-Digit SIC: Z_dWA and Z_dWS . . . 189

LIST OF FIGURES

2.1 Distribution of Fuzz Factors 64

2.2 Distribution of Noise: B 85

2.3 Distribution of Noise: W_1 86

Chapter 1

Wage Dispersion, Compensation Policy, and the Role of Firms [*]

1.1 Introduction

Empirical work in economics stresses the importance of unobserved firm- and person-level characteristics in the determination of wages, finding that these unobserved components account for the overwhelming majority of variation in wages. However, little is known about the mechanisms sustaining these wage differentials. This paper attempts to demystify the firm-side of the puzzle by developing a statistical model that enriches the role that firms play in wage determination, allowing firms to influence both average wages as well as the returns to observed worker characteristics.

A burgeoning literature exists focused on measuring employer wage differentials. In nearly all of these studies, firm wage differentials are measured as a firm-specific intercept: Groshen (1991) establishes the importance of establishment

[*]This chapter reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureaus Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical, methodological, and technical issues are those of the author[s] and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov).

wage differentials relative the inter-industry wage differentials; Groshen and Levine (1998) measure the magnitude and persistence of firm- and occupation-wage differentials in an attempt to assess the importance of internal labor markets; Bronars and Famulari (1997) provide evidence of employer wage differentials in their investigation of the determinants of wage levels and wage growth; while Abowd et al. (1999) use employer-employee connected data to implement a technique for estimating a wage model with both firm and person effects. These differentials are often decomposed into the portion explained by observed firm characteristics, for example, industry of operation, and an unobserved component – the portion of wages explained by firm identity that is not explained by firm characteristics available to the econometrician. A similar decomposition is applied to the individual component of wages. Person-level intercepts – the time invariant portion of wages attributed to individuals – are decomposed into a component explained by time invariant, though observed, characteristics such as race or sex and a component that is unobserved. The person-specific unobserved portion of wages is often thought of as reflecting the value of an individual's innate ability or talent that is portable across firms.

A complementary literature explores the relationship between firm characteristics and wage outcomes. For example, empirical work suggests that there exists a significant relationship between wages and firm size. Brown and Medoff (1989) provides an analysis of the role of firm size in the determination of wages. Large firms may pay higher wages because large firms hire higher quality workers, offer inferior working conditions, are more threatened by unionization, or face higher monitoring costs (*e.g.*, efficiency wages). Davis and Haltiwanger (1995), using data from the Current Population Survey and Census of Manufacturers, investigate the relationship between firm size and both wage levels and wage dispersion. They find that average wages are higher at larger establishments and that wage dispersion is inversely related to firm size class. In addition, their work suggests that the returns to observed characteristics vary across firm size class.

Finally, economists are also looking within the firm to explain wage setting from the perspective of compensation and strategic management policies. As Lazear (2000) mentions in his review of the personnel economics literature, firms may influence wage determination and wage dynamics through firm level policies by designing deferred compensation policies to motivate workers, tying remuneration to observed productivity, or designing tournament schemes. Firms may also pay efficiency wages to dissuade shirking and may provide insurance against shocks to the value of labor services (Prendergast 1999). Recent work in Ichniowski and Shaw (2003) and Cappelli and Neumark (2001) evaluates the relationship between firm-level human resource practices and firm productivity and wage outcomes. Finally, the work of Baker et al. (1995) illustrates the value in evaluating the wage policies of an individual firm.

The empirical work focused on decomposing wages, emphasizing the relative importance of the returns to observed characteristics and unobserved heterogeneity, has in some sense preceded the evolution of economic theory that explains why unobserved heterogeneity exists. Firms pay vastly different wages, though observed firm characteristics fall short of explaining the overwhelming majority of the variation in average wages across firms. As economists begin to look inside the firm for evidence supporting this measured heterogeneity, they are finding that firms pursue idiosyncratic compensation policies and human resource management strategies. The work in this paper pushes this investigation one step further, suggesting that digging deeper into the black box of firm compensation policy requires evaluating not only traditional firm- and person-specific determinants of wages, but also the way in which the characteristics of workers are valued by firms. Heterogeneity in firms' policies, which may explain why firms pay different average wages, may also suggest that certain worker attributes are more valuable to particular firms.

Using employer-employee linked data for the United States, I specify and estimate a multilevel statistical model of earnings determination that measures variation in firm wage policies. The specification controls for person-level, unobserved heterogeneity and allows for firm-specific deviations in average firm wages as well as in the returns to components of human capital: race, gender, education, and experience. The estimation procedure also provides estimates of the elements of the variance matrix of pay policy parameters. Results suggest that there is statistically significant variation in the returns to worker characteristics across firms. For instance, the estimated variance of the returns to being non-white, for both men and women, is roughly the same size as the variance of average wages across firms. Allowing for an unstructured covariance matrix of firm-specific returns reveals that firms that tend to pay higher average wages also tend to pay higher than average returns to certain worker characteristics. Finally, roughly 7% of wage variation is accounted for by the firm-specific returns to human capital. Though it appears that the majority of this explained variation would have otherwise – in the absence of firm-specific returns – been accounted for by person- and firm-specific intercepts, firm-specific valuation of human capital explains a greater proportion of wage variation than the returns to the observed firm characteristics included in the estimation: the level of firm employment and industry division of operation.

Hierarchical modeling techniques are used in Cardoso (2000) to analyze the relationship between firm level characteristics and worker wages in a cross-section of employer-employee connected data for Portugal. The model presented here extends the approach in Cardoso (2000) by integrating it into a larger literature that accounts for unobserved heterogeneity across workers and firms, by permitting a fully unstructured variance matrix of pay policy parameters, and by accounting for the total variation in wages attributed to the firm-specific returns to human capital. My model may also be viewed as an extension of Abowd et al. (1999) in that it allows unobserved heterogeneity across firms to be captured by firm-specific intercepts as well as firm-specific returns. By exploiting the longitudinal and connected nature of the employer-employee dataset, I am able to control for both unobserved firm- and person-effects. I use Restricted Maximum Likelihood (REML) techniques to estimate the model variance and covariance parameters and derive random coefficient estimates using the Henderson (or Mixed) Model equations. Unlike traditional approaches to estimating random effects models, for instance, Generalized Least Squares, the REML approach does not assume orthogonality between the random effects – the firm-specific slopes and firm- and person-specific intercepts – and observed covariates. My approach also allows for a fully unstructured variance matrix of the firm-specific intercepts and returns. Finally, I provide a decomposition of the total variation in wages attributed to both the observed and unobserved components in the model, thereby highlighting the relative importance of the firm-specific components of compensation.

The paper is organized as follows. Section 1.2 discusses the multilevel model in the context of firm compensation policy; Section 1.3 develops the statistical model and discusses the estimation technique; Section 1.4 provides a description of the data; Section 1.5 provides results; and Section 1.6 concludes.

1.2 Firm Compensation Policy

To evaluate the influence of firms on wages, I specify a multilevel (mixed) model of wage determination. The multilevel characteristic of the model captures the inherently hierarchical nature of the data: individuals hold jobs (level 1) that are nested within firms (level 2). The model is mixed in that it contains both random and fixed coefficients. Firms shape individual wage outcomes in two ways: through firm-specific average wage effects and firm-specific price effects. Average wages are influenced by firm-specific intercepts as well as by observed characteristics, namely, firm employment (size) and industry division of operation. Firms also influence wage setting through the returns to the characteristics of human capital: sex, race, education, and experience.[1] These returns have an unobserved firm-specific component, measured as deviations around the sample average return for a given characteristic of human capital, as well as an observed component – the

[1] In the implementation, the model is fully interacted by sex; firm-specific returns are estimated for both men and women.

interaction between observed characteristics of workers and firms. For example, consider the influence of firms on wage setting when only firm size and experience are observed. Firms will influence wages through a firm-specific average wage effect that has two components: an unobserved, though identifiable, firm-specific random intercept and the observed effect of firm size. Firms will also influence wages through the returns to experience, which also has two components: an unobserved, though identifiable, firm-specific random return to experience and the observed interaction between experience and firm size.

The statistical model retains characteristics of more standard models of wage determination. Workers earn average (or market) returns to the components of human capital which are measured as fixed coefficients on the returns to human capital. Thus, the firm-specific random returns may be viewed as deviations around the sample average return to the components of human capital. A person-specific, random intercept is also introduced to control for unobserved heterogeneity and is usually interpreted as the return to a worker's innate ability or skill. Finally, a number of dummy variables are included to control for time effects and issues related to the construction of the data.[2]

The econometric specification, while driven partly by the structure of the data, is loosely supported by theoretical work in two recent papers. In Abowd et al. (2006), a simple model of production, wage determination, and mobility is specified. There are two types of firms: complex and simple. In simple firms,

[2]These variables are discussed in further detail in Section 4. A list of variables included in the estimation is presented in Table 1.1.

wages are set equal to a worker's productivity which is determined by her innate ability and level of experience. Productivity in complex firms, however, is firm-specific and is a function of an individual's innate ability and firm-specific seniority. Wages are determined by a simple sharing rule and, for workers at complex firms, wages are thus a function of tenure at the firm. The underlying model supports an empirical specification in which a firm-specific wage policy captures both the sharing rule and a production technology that is a function of tenure at the firm. The model, while particularly relevant to the literature on mobility and the returns to tenure, suggests a broader treatment of firm specificity in production and the formation of compensation policy. Lazear (2003) provides another simple and more general model of output and wage determination. A worker's output is determined by a set of general skills which are valued by all firms in the economy. Firms differ, however, in the relative weighting of these skills in production. Workers and firms split the value of output and a worker's earnings are determined by his share of output – a function of his optimal investment in general skills and the firm's specific weighting of his skills. The model is primarily developed to provide an explanation for the empirically observed returns to firm-specific tenure, though is relevant to the statistical model presented in this paper which allows for firm-specific returns to the components of human capital that are used by all firms. In light of the model developed in Lazear (2003), the firm-specific returns to human capital in this paper may be viewed as capturing the firm-specific weighting of these characteristics in production.

Efficiency wage theory, which generally refers to extra-marginal wage payments at the firm-level, may also support a model of firm-specific returns. Firms that pay high average wages to all of their workers may do so to provide a disincentive to shirking. It may be, however, that the level of certain worker attributes – experience and education – make monitoring by firms more difficult. More experienced workers may be more likely to shirk because, over their careers, they have learned to shirk successfully. It may also be that workers with more experience and education are more likely to have more complex jobs where output is hard to observe and, thus, hard to monitor. In either case, one might expect to see higher than average firm-level returns to these characteristics where monitoring is difficult. If the existence of high average firm wages suggests that firms are paying efficiency wages, then it seems reasonable – if the monitoring of highly educated and experienced workers is more difficult – to expect a positive covariance between firm-specific average wages (intercepts) and firm-specific returns to education and experience.

Regardless of theoretical motivation for the multilevel specification, the structure of the model assumes that worker-firm attachments, *i.e.*, the clustering of workers within firms, is meaningful. The treatment of firm-specific deviations as random effects provides an efficiency gain over other modeling approaches that do not take into account the clustering of workers within firms in cases where this clustering is present in the data.[3] A richer error specification will provide correct standard errors for the fixed effects in the model. Random firm-specific effects identified in the data, however, may exist for a variety of reasons. For instance, consider the specific return to education. If the measure of education is not quality adjusted,

[3] An example is Ordinary Least Squares (OLS), where the model error is assumed spherical.

which is the case in the data used in this analysis, a higher firm-specific return may be due to the clustering of workers with high levels of education from high quality universities who earn high wages relative to workers at the same firm who have lower levels of education from low quality schools. Dispersion in the returns to education, thus, would result from the inability to observe the quality of education. Accounting for this clustering is meaningful in the aforementioned sense. However, if the force driving dispersion in the returns to education is due to unobserved quality issues, the economic content of these slopes is questionable. It may be, however, that dispersion in returns across firms are meaningful in economic terms. In the previous example, if firms that hire the highest quality university graduates also hire the highest quality high school graduates, then the existence of higher than average firm-level returns to education may reflect a hiring policy: always hire the highest quality of graduates regardless of level of education. As will be made clear in the development of the model, these firm-specific deviations may be thought of as the firm-specific unobserved error components in a pricing model of the characteristics of human capital. From this perspective, the existence of firm-specific returns is due to the inability of the observed firm characteristics in the model to explain variation in the returns to the characteristics of human capital. In either case, the approach permits an accounting for the contribution of observed and unobserved person and firm characteristics to variation in wages.

In addition to allowing for firm-specific intercepts and returns, the multilevel model is flexible in accounting for both the variance and covariance of these firm-specific returns. For each firm-specific component, a sample-wide variance is specified.[4] The structure of the covariance of these effects, however, may take a variety of forms. In this analysis, I evaluate both a diagonal and fully unstructured variance matrix of firm effects. Exploring the unstructured form of the of this matrix addresses potentially important questions regarding firm wage policy. For instance, do firms that tend to pay high wages on average also tend to reward more highly the returns to education? Is the dispersion in the returns to education across firms similar for men and women?

The model also includes a person-specific random intercept. Though jobs are nested within firms, the person-specific random intercept exists outside the hierarchy that is developed in the following section. It is included to capture some of the unobserved heterogeneity in wages that is due to what is traditionally considered an individual's innate ability or skill and also to serve as a basis for comparison to work that models heterogeneity in wages with both firm- and person-specific intercepts. Future work will attempt to integrate more fully the person-specific intercepts into the variance structure of the multilevel model. For instance, permitting the covariance between the person-specific intercept and the firm-specific returns to education would provide a direct empirical test of whether individuals with high levels of innate ability (higher than average person-intercepts) are more likely to earn higher returns to experience.

A thorough discussion of multilevel models is provided in Goldstein (1995). Raudenbush and Bryk (1986) also provide a discussion and application to estimating

[4]Each effect is assumed to have a zero mean. In the analysis, the data are de-meaned so that dispersion of the firm intercepts is appropriately centered. For the firm-specific slopes, the fixed coefficient estimates of the returns to the components of human capital provide estimates of the mean of each distribution.

the returns to student achievement. Cardoso (2000) and Cardoso (1999) develop a multilevel wage model using employer-employee connected data for Portugal that is estimated using Iterative Generalized Least Squares (IGLS) techniques that are popular in the multilevel model literature. The multilevel model, which includes both fixed and random coefficients, is also essentially a mixed model. Estimation of and prediction in mixed models is discussed at length in McCulloch and Searle (2001) and Searle et al. (1992). The approach in this paper draws heavily on these techniques.

A general multilevel model is outlined in the following section and is shown to have a mixed model representation. Also discussed is the three-part estimation approach: (1) the natural log of the real wage, the dependent variable, is transformed, removing variation attributed to the variables for which coefficients are fixed; (2) Restricted Maximum Likelihood (REML) techniques are used to estimate the variance components of the model; and (3) predictors of the model's fixed and random effects are derived using the Henderson (or Mixed Model) equations. Attention is given to issues surrounding the identification of the model's random effects and issues involved in the identification of these effects if they were assumed fixed.

1.3 Statistical Model

1.3.1 Multilevel Model

Let $i = 1, ..., N$ index workers, $j = 1, ..., J$ index firms, and $t = 1, ..., T$ index time. The natural log of the real wage, w_{ijt}, for worker i employed by firm j in time t is modeled as:

$$w_{ijt} = x_{ijt}^{(1)\prime}\beta^{(1)} + x_{it}^{(2)\prime}\gamma_{jt} + \alpha_i + \varepsilon_{ijt} \tag{1.1}$$

$$x_{ijt}^{(1)} = \begin{bmatrix} x_{1ijt}^{(1)} \\ \vdots \\ x_{mijt}^{(1)} \end{bmatrix}, x_{ijt}^{(2)} = \begin{bmatrix} x_{1ijt}^{(2)} \\ \vdots \\ x_{kijt}^{(2)} \end{bmatrix}$$

where $x_{ijt}^{(1)}$ is an $(m \times 1)$ vector of person and firm varying covariates for which parameters $\beta^{(1)}$an $(m \times 1)$ vector are fixed;[5] $x_{it}^{(2)}$ is a $(k \times 1)$ vector of person-level components of human capital (including an intercept) for which parameters γ_{jt} a $(k \times 1)$ vector are firm-specific;[6] $\alpha_i \sim N(0, \sigma_\alpha^2)$ is a person-level random intercept; and $\varepsilon_{ijt} \sim N(0, \sigma_\varepsilon^2)$ is a residual error term.

The second level of the model involves specifying a relationship between γ_{jt} – the vector of firm-specific returns to human capital – and firm characteristics:

$$\gamma_{jt} = g_{jt}'\beta^{(2)} + \Psi_j \tag{1.2}$$

where:

[5] The elements of $x_{ijt}^{(1)}$ are: a race missing dummy, negative experience dummy, and time effects interacted with sex. The superscript (1) signifies level 1 covariates for which parameters do not vary over the second level.

[6] The elements of $x_{it}^{(2)}$ include: a constant, experience, $\frac{experience^2}{100}$, education, and a nonwhite dummy, all of which are interacted with sex. The superscript (2) is to signify level 1 covariates for which parameters do vary over the second level.

$$\underset{(kl\ x\ k)}{g_{jt}} = I_k \otimes f_{jt} \tag{1.3}$$

f_{jt} is an (l ×1) vector of time-varying firm covariates – firm-level employment and industry of operation:

$$f_{jt} = \begin{bmatrix} f_{1jt} \\ \vdots \\ f_{ljt} \end{bmatrix} \text{ where } f_{1j} = 1, \Psi_j = \begin{bmatrix} \psi_{1j} \\ \vdots \\ \psi_{kj} \end{bmatrix} \tag{1.4}$$

$\beta^{(2)}$ is a (kl ×1) vector of fixed parameters describing the relationship between these firm characteristics and the returns to person-level covariates and average wages. Ψ_j is a (k ×1) vector of firm-specific random errors with each element of Ψ_j corresponding to each of the k components of human capital. Thus, firms influence the returns to human capital in two ways. The term $g'_{jt}\beta^{(2)}$ reflects the influence of time varying, observed firm characteristics. For example, the firm-specific returns to education, in the context of this model, are influenced by both the size and industrial classification of the employing firm. This component will capture whether firms that tend to be large in terms of employment also tend to pay higher returns to experience. The vector Ψ_j, the firm-specific error component, captures the influence of the unobserved, firm-specific component of compensation policies. All variation across firms in the return to a component of human capital that is not captured by $g'_{jt}\beta^{(2)}$ is accounted for by the elements of Ψ_j. This error representation, for example, would capture the existence of firm-specific human resource or management policies that would otherwise not be captured by firm observed characteristics. Finally, it is important to notice that while the firm-specific parameters γ_{jt} vary over time, this variation is due to time variation in the value of f_{jt} and not Ψ_j. Thus, stochastic changes in the firm-specific error components are smoothed over time.[7]

A characteristic of the model is that the vector of firm-specific errors Ψ_j is permitted to have a fully unstructured variance matrix. It is assumed that $\Psi_j \sim N(0, \Gamma)$ where:

$$\Gamma = \begin{bmatrix} \sigma^2_{\psi_1} & \cdots & \sigma_{\psi_1\psi_k} \\ \vdots & \ddots & \vdots \\ \sigma_{\psi_1\psi_k} & \cdots & \sigma^2_{\psi_k} \end{bmatrix} \tag{1.5}$$

The characterization of Γ in equation (1.5) is general, capturing the variances of the firm-specific intercepts and returns *as well as* the covariances between these terms. For instance, $\sigma^2_{\psi_1}$ is the variance of the sample-wide firm-specific intercepts, measuring the dispersion in wages due to dispersion in firm-specific average wages. The extent to which firms in sample tend to pay high average wages and high returns to education, experience, and race for men and women will be captured by covariance terms in the first column (or first row) of Γ.

Combining equations (1.1) and (1.2) yields:

[7] This will turn out to be important as the components of Ψ_j will later be interpreted as firm-specific returns to human capital. Increases in the returns to these components – for example, due to increases in the returns to skill – will be smoothed over the 1990-1998 sample used in the analysis.

$$w_{ijt} = x_{ijt}^{(1)\prime}\beta^{(1)} + x_{it}^{(2)\prime}g_{jt}'\beta^{(2)} + x_{it}^{(2)\prime}\Psi_j + \alpha_i + \varepsilon_{ijt} \tag{1.6}$$

The influence of the firms (second level of the model) is readily seen in equation (1.6). The first term on the right hand side of equation (1.6) captures the influence of the person-level covariates for which parameter variation is not firm-specific – the set of time and control variables. The second term captures the influence of observed, firm-level characteristics (influencing variation in γ_{jt}) on the components of human capital. Recalling the structure of g_{jt} in equation (1.2), the multiplication of $x_{it}^{(2)\prime}g_{jt}'$ yields a vector containing the components of human capital, firm observed characteristics, and the interactions between the components of human capital and observed firm characteristics. The returns to the components of human capital in $x_{it}^{(2)\prime}g_{jt}'\beta^{(2)}$ are viewed as the market or sample-wide average returns to these characteristics; the returns to observed firm characteristics and the interactions capture the influence of the observed firm characteristics in equation (1.2). The firm-specific random errors Ψ_j in equation (1.2), in the context of equation (1.6), become a set of firm-specific random coefficients for the components of human capital.

In order to explain the technique used to estimate the model, it is helpful to express the model in matrix notation. Doing this takes a bit of finessing. Express $x_{ijt}^{(1)\prime}\beta^{(1)} + x_{it}^{(2)\prime}g_{jt}'\beta^{(2)}$ in (1.6) as:

$$x_{ijt}^{(1)\prime}\beta^{(1)} + x_{it}^{(2)\prime}g_{jt}'\beta^{(2)} = b_1^{(1)}x_{1ijt}^{(1)} + \cdots + b_m^{(1)}x_{mijt}^{(1)} + b_1^{(2)}x_{1it}^{(2)}g_{1jt} + \cdots + b_{kl}^{(2)}x_{kit}^{(2)}g_{ljt} = x_{ijt}'\beta$$

where x_{ijt} is a $((kl+m)\ \times 1)$ vector of covariates – those contained in $x^{(1)}_{ijt}$ which are specified as having fixed coefficient estimates in equation (1.1) and the interactions of the $x^{(2)}_{ijt}$ in equation (1.1) with the firm-level characteristics f_{jt} in equation (1.4). β is a $((kl+m)\ \times 1)$ vector of fixed coefficients. Equation (1.6) becomes:

$$w_{ijt} = x'_{ijt}\beta + x^{(2)\prime}_{it}\Psi_j + \alpha_i + \varepsilon_{ijt} \tag{1.7}$$

Let N index the number of observations, I index the number of persons, and J index firms. Grouping observations into firms and stacking yields:

$$w = X\beta + X^{(2)}\Psi + D\alpha + \varepsilon \tag{1.8}$$

where w is an $(N \times 1)$ vector of earnings; X is an $(N \times (kl+m))$ matrix of covariates; and β is a $((kl+m) \times 1)$ vector of fixed coefficients. $X^{(2)}$ is an $(N \times Jk)$ stacked matrix of person-level characteristics (the $x^{(2)\prime}_{it}$ grouped by firm):

$$X^{(2)} = \begin{bmatrix} X^{(2)}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & X^{(2)}_J \end{bmatrix} \tag{1.9}$$

so that $X^{(2)}_j$ is an $(n_j \times k)$ matrix of observations of workers at firm J. Ψ is a $(Jk \times 1)$ vector of firm specific random returns:

$$\Psi = \begin{bmatrix} \Psi_1 \\ \vdots \\ \Psi_J \end{bmatrix} \tag{1.10}$$

and is ordered so that the firm specific returns in Ψ_j (recall that Ψ_j is a (k $\times 1$) vector) correspond appropriately to the block of observations in $X^{(2)}$ for that firm. Formally, $X^{(2)}$ is the design of firm-specific random effects and includes firm-specific random intercepts and returns to worker level characteristics. Finally, D is an $(N \times I)$ design matrix for the random person effects contained in the $(I \times 1)$ vector α and ε is an $(N \times 1)$ vector of residuals.

The model's error is represented in matrix notation as follows:

$$\begin{bmatrix} \Psi \\ \alpha \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_J \otimes \Gamma & 0 & 0 \\ 0 & \sigma^2_\alpha I_I & 0 \\ 0 & 0 & \sigma^2_\varepsilon I_N \end{bmatrix} \right) \tag{1.11}$$

While the variance of the firm-specific returns and intercepts is fully unstructured – the term $I_J \otimes \Gamma$ – and a structure is placed on the person-specific intercepts $\sigma^2_\alpha I_I$, the person-specific intercepts α are not permitted to co-vary with the elements of Ψ. This is a restrictive assumption. While the thrust of this paper is to explore a specific characterization of firm compensation policy – namely, one in which firms are permitted to pay specific returns and to adjust these parameters such that high returns to some components may be associated with high or low returns to others – it is reasonable to believe that individuals with higher than average levels of innate ability, measured by relatively high α's, may also earn higher returns to certain characteristics or may be more likely to match to firms that pay higher than average wages. A context in which the current error structure in (1.11) seems reasonable is one where firms are unable to implement compensation policies conditional on innate (and initially unobserved) worker ability and where learning about innate ability, through employment at the firm, does not induce adjustments in firm compensation policy.

1.3.2 Model Estimation and Prediction

Methods for estimating the model specified by equations (1.8) and (1.11), in the context of the multilevel model literature, are discussed in Goldstein (1995) and include Iterative Generalized Least Squares (IGLS), Bayes or Empirical Bayes, and Generalized Estimating Equations (GEE) methods. Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) are briefly mentioned as well. Estimation of the model has also received considerable treatment in the statistics literature. Equations (1.8) and (1.11) represent a mixed model formulation as the specification contains both random and fixed coefficients.

The specifications in this paper are estimated using REML, which provides estimates of the variance components in (1.11): $\widehat{\sigma}^2_\alpha$, $\widehat{\sigma}^2_\varepsilon$, and the elements of $\widehat{\Gamma}$. Loosely speaking, "a basic idea of restricted maximum likelihood (REML) estimation is that of estimating variance components based on residuals calculated after fitting by ordinary least squares just the fixed effects part of the model" rather than basing these estimates on the dependent variable (Searle et al. (1992) p. 250). See Appendix A for a more detailed discussion of the REML technique.

Using the variance estimates emerging from REML, predictors of the random coefficients – $\widehat{\Psi}$ and $\widehat{\alpha}$ – and estimates of the fixed coefficients $\widehat{\beta}$ in equation (1.8) are derived from the Mixed Model (or Henderson) Equations. See Appendix B for a full derivation of the predictors and estimates as well as a discussion of their identification.

1.4 Data Description

I estimate the multilevel model using data that are housed at the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program. A thorough discussion of the data maintained by LEHD is provided in Abowd et al. (2005); a brief description is provided here. The LEHD Program maintains a variety of survey and administrative data from a number of federal and state agencies. State unemployment insurance (UI) wage records and Quarterly Census of Employment and Wages (QCEW) establishment data are available for the states in partnership with the LEHD Program.[8] State-level unemployment insurance (UI) data contain quarterly earnings for employees covered by state unemployment insurance systems

[8] Over 40 states have signed Memoranda of Understanding to engage in data sharing with the U.S. Census Bureau. For 34 of these states, core infrastructure and public use data are available. More information on state partnerships with the U.S. Census Bureau's LEHD Program is available at: http://lehd.dsd.census.gov.

(over 96% of private sector employment is covered by the UI universe) as well as both a person identifier – a person-specific Protected Identification Key (PIK) – and a firm identifier – a State Employer Identification Number (SEIN). The structure of the UI data conforms to the data requirements of the multilevel model specifications: there are multiple workers in the UI data for whom wages are reported within the same firm (workers are clustered within firms). Thus, a firm, as defined in this analysis, is a collection of workers who share a common SEIN.[9] And, individual wage records are then linked across quarters on the basis of the PIK to create individual work histories.

Nearly all of the firm identifiers in the UI data are found in the universe of state QCEW data which contain richer information on firms. However, while groupings of workers by a common SEIN define the smallest firm in the UI data, a finer definition of the firm is available in the QCEW data.[10] For SEINs in the QCEW data that operate more than one establishment, firm characteristics such as employment, total payroll, location, and industry of operation are reported at the level of the establishment. While the finer, establishment-level detail is not used in this analysis, the QCEW do provide the measure of industry of operation. For firms in the QCEW reporting under multiple establishments (a multi-unit firm), the SEIN-level industry is the employment weighted modal industry of the underlying establishments. For firms reporting under a single establishment, the reported industry is used.

[9] The SEIN is by definition a state-level firm identifier; thus, firms are also state-specific.

[10] This is true for all states except Minnesota. The Minnesota UI data provide two firm identifiers – an SEIN and reporting unit number – which are consistent with the level of reporting in the QCEW.

Finally, worker demographics – sex, date of birth, and race – are acquired from the Census NUMIDENT file (a version of Social Security Administration person-level micro-data) and are matched to the UI wage record data on the basis of the person identifier. Neither the UI nor the Census NUMIDENT data provide a direct measure of education for individuals in the sample. Education information is available, however, in the Decennial Census of Population; in this analysis, the education variable for workers in the UI data is based on a statistical match to the 1990 Decennial Census of Population.[11] The race variable is collapsed into an indicator for non-white, rather than used directly in the model estimation.[12]

All worker and firm measures are transformed into annual values. From the QCEW data, only industry of operation is retained in this analysis. For each firm (SEIN), the annual (calendar year) employment weighted, quarterly modal industry is used.[13] The measure of firm size is simply the summation of workers with positive earnings at the firm within the calendar year. Date of birth from the NUMIDENT is used to create an age measure which, in turn, is used to create a measure of potential experience. For the first year that an individual appears in the UI data, potential experience is calculated as the person's age at the beginning of the quarter minus her years of education minus 6. This potential experience measure is then

[11]Work is currently being done at LEHD to match workers in the UI wage data to the 2000 Decennial Census of Population on the basis of PIK, thus, providing reported education measures for roughly 1 in 6 workers in the UI universe. For remaining workers in the UI data, education levels will be created using multiple imputation techniques.

[12]The collapsing of the race variable is common in nearly all of the wage research using LEHD data. Preliminary research into the quality of the race variable on the NUMIDENT raised concerns regarding it's usefulness in analysis. In the future, I plan to explore the reliability of using the race variable in the context of the model developed in this paper.

[13]Firm data in the QCEW are reported on a quarterly basis. While it is rare for firms to change industry classification, the algorithm ensures that the most important industry – in terms of employment – is used in the analysis.

augmented by observed years of experience: years during which the individual has positive earnings in the UI data.

The wage measure used in the analysis is based on an annualization of the quarterly payroll values reported in the UI data. For workers in the UI data who hold multiple jobs concurrently, the dominant job with the highest level of annual earnings is retained. Within each dominant job, quarterly earnings within the calendar year are used to create a measure of annualized earnings (wages). The sequence of earnings within a job are used to categorize workers into three groups which are then used to construct the annual earnings measure:

- Annualized earnings based on full-quarter status. A full-quarter worker is one who has positive earnings at a firm in the current (t), previous ($t-1$), and subsequent ($t+1$) quarters at the same firm. For workers who have worked at least one full quarter at a firm, 4 times the full-quarter average of earnings is used to construct the annual measure. Roughly 84% of the annual earnings are constructed this way.

- Annualized earnings based on continuous-quarter status. A continuous-quarter worker has positive earnings at a firm in the current (t) and previous ($t-1$) quarters or the current (t) and subsequent ($t+1$) quarters at the same firm. For workers who have not worked one full quarter during the calendar year, but who have at least one continuous quarter of employment, 8 times the average of continuous-quarter average earnings is used to construct the annual measure.[14]

- Annualized earnings based on reported quarterly earnings. For workers who are neither classified as full- or continuous-quarter for at least one quarter in the calendar year, 12 times the average of quarterly earnings is used to construct the annual measure.

For all observations, dummy variables are created to control for the type of annualized earnings measure that is used in the estimation of the wage model. The analysis sample is restricted to include employees who are between the ages of 25 and 65, have real annualized earnings between $1,000 and $1,000,000, and who work at firms with more than 10 employees that are not operating in agriculture or public administration Standard Industrial Classification (SIC) industry divisions.

Estimation of the multilevel model using data for all states providing data to the LEHD program is a computationally infeasible task. In order to estimate the model, I first select three states and then draw a sample of workers from these states.[15] A requirement of the multilevel model is that workers are clustered within firms, thus, a procedure for sampling must guarantee that workers are clustered within firms. The sampling procedure developed in Woodcock (2003) is used and guarantees that the sample is representative of employment and that a minimum number of workers – here, at least 10 – are sampled from each firm. For 1997,

[14]The assumption is that a continuous-quarter worker (who is also not a full-quarter worker) has an expected employment duration of 0.50. Thus, observed continuous-quarter earnings are, in expectation, 50% of unobserved full-quarter earnings.

[15]The identities of the states used in the analysis are witheld for confidentiality reasons. The three states chosen, however, are geographically dispersed across the United States and contain both large urban and rural areas.

a sample of firms is drawn with probabilities that are proportional to firm-level employment. Workers within those firms are sampled with probabilities that are inversely proportional to the firms' employment. Finally, the entire earnings histories for the sampled worker-firm pairs are included in the final dataset. The resulting datafile is a random sample of roughly 0.25% of workers in the three states and contains 283,507 annual observations on 55,267 individuals and 29,591 firms for the 1990-1998 time period.

Table 1.1 summarizes the list of variables used in the analysis, including a brief description of each. Tables 1.2 and 1.3 present means and standard deviations of relevant characteristics for the analysis sample.

1.5 Estimation and Results

1.5.1 Specifications

I present estimates for four versions of the multilevel model. Each successive specification is a generalization of the preceding specification, reflecting the increasingly richer characterization of the firm-specific returns to human capital γ_{jt} and the variance matrix of the random error components in Γ. Specification 1 is a wage decomposition that allows for random person- and firm-intercepts and fixed returns to the observed components of human capital. This specification serves as the baseline wage model in which heterogeneity is captured only through person- and firm-specific intercepts. Specification 2 builds on Specification 1 by introducing the firm-specific returns to human capital; moreover, the variance matrix of these

Table 1.1: Variables Used in Analysis

Demographics	
Education	Based on statistical match to Decennial Census 1990
Non-white	Based on race variable in Census NUMIDENT
Race Missing	Based on race variable in Census NUMIDENT
Sex	Based on sex variable in Census NUMIDENT
Job Characteristics	
ln(Annualized Real Wage)	Annualized wage measure based on quarterly earnings (UI)
Experience	Potential experience measure constructed using observed experience and date of first appearance in sample
Age	Base on date-of-birth measure reported in the Census NUMIDENT
Negative Experience Dummy	Dummy for negative values of potential experience
Firm Characteristics	
ln(Firm Employment)	Natural log of the sum of workers with positive annualized earnings
SIC 2-Digit Group (10-17)	Mining and Construction
SIC 2-Digit Group (20-29)	Manufacturing
SIC 2-Digit Group (30-39)	Manufacturing
SIC 2-Digit Group (40-49)	Transportation, Communications, Electric, Gas, and Sanitary Services
SIC 2-Digit Group (50-59)	Wholesale and Retail Trade
SIC 2-Digit Group (60-67)	Finance, Insurance, and Real Estate
SIC 2-Digit Group (70-79)	Services
SIC 2-Digit Group (80-89)	Professional Services
Time Dummies	
4 Full Quarters Worked 1990	Dummy
4 Full Quarters Worked 1991	Dummy
4 Full Quarters Worked 1992	Dummy
4 Full Quarters Worked 1993	Dummy
4 Full Quarters Worked 1994	Dummy
4 Full Quarters Worked 1995	Dummy
4 Full Quarters Worked 1996	Dummy
4 Full Quarters Worked 1997	Dummy
4 Full Quarters Worked 1998	Dummy
Less Than 4 Full Quarters Worked 1990	Dummy
Less Than 4 Full Quarters Worked 1991	Dummy
Less Than 4 Full Quarters Worked 1992	Dummy
Less Than 4 Full Quarters Worked 1993	Dummy
Less Than 4 Full Quarters Worked 1994	Dummy
Less Than 4 Full Quarters Worked 1995	Dummy
Less Than 4 Full Quarters Worked 1996	Dummy
Less Than 4 Full Quarters Worked 1997	Dummy
Less Than 4 Full Quarters Worked 1998	Dummy
Discontinuous Employment Dummy	Dummy
0 Full Quarters Worked	Dummy
1 Full Quarters Worked	Dummy
2 Full Quarters Worked	Dummy
3 Full Quarters Worked	Dummy
4 Full Quarters Worked	Dummy
State Dummies	
State 1	Dummy
State 2	Dummy
State 3	Dummy

Table 1.2: Sample Demographics and Job Characteristics

	Men		Women	
	Mean	Standard Deviation	Mean	Standard Deviation
Demographics				
Proportion Male	0.5290	0.4992		
Education	12.5484	2.9799	12.7728	2.5678
Proportion Non-white	0.3216	0.4671	0.3335	0.4715
Proportion Race Missing	0.0422	0.2011	0.0321	0.1763
Job Characteristics				
ln(Annualized Real Wage)	10.6035	0.7441	10.2452	0.6706
Experience	21.9814	9.9102	21.9719	9.8157
Age	40.1197	9.5078	40.2895	9.3661
ln(Firm Employment)	6.1190	2.3237	6.5018	2.2787
Proportion of Employment:				
SIC 2-Digit Group (10-17)	0.0816	0.2738	0.0210	0.1434
SIC 2-Digit Group (20-29)	0.0782	0.2685	0.0608	0.2390
SIC 2-Digit Group (30-39)	0.1799	0.3841	0.0906	0.2870
SIC 2-Digit Group (40-49)	0.1067	0.3087	0.0676	0.2511
SIC 2-Digit Group (50-59)	0.2125	0.4091	0.1752	0.3802
SIC 2-Digit Group (60-67)	0.0614	0.2401	0.1062	0.3081
SIC 2-Digit Group (70-79)	0.1091	0.3118	0.0880	0.2833
SIC 2-Digit Group (80-89)	0.1706	0.3761	0.3906	0.4879
4 Full Quarters Worked 1990	0.0461	0.2096	0.0478	0.2134
4 Full Quarters Worked 1991	0.0483	0.2144	0.0498	0.2176
4 Full Quarters Worked 1992	0.0623	0.2418	0.0629	0.2429
4 Full Quarters Worked 1993	0.0728	0.2598	0.0756	0.2643
4 Full Quarters Worked 1994	0.0897	0.2857	0.0915	0.2883
4 Full Quarters Worked 1995	0.1020	0.3027	0.1047	0.3062
4 Full Quarters Worked 1996	0.1252	0.3309	0.1295	0.3358
4 Full Quarters Worked 1997	0.1363	0.3431	0.1392	0.3462
4 Full Quarters Worked 1998	0.1257	0.3315	0.1292	0.3354
Less Than 4 Full Quarters Worked 1990	0.0122	0.1098	0.0116	0.1070
Less Than 4 Full Quarters Worked 1991	0.0125	0.1112	0.0123	0.1101
Less Than 4 Full Quarters Worked 1992	0.0158	0.1246	0.0163	0.1265
Less Than 4 Full Quarters Worked 1993	0.0179	0.1326	0.0168	0.1287
Less Than 4 Full Quarters Worked 1994	0.0195	0.1381	0.0177	0.1317
Less Than 4 Full Quarters Worked 1995	0.0273	0.1631	0.0246	0.1549
Less Than 4 Full Quarters Worked 1996	0.0376	0.1903	0.0345	0.1824
Less Than 4 Full Quarters Worked 1997	0.0808	0.2725	0.0714	0.2574
Less Than 4 Full Quarters Worked 1998	0.0408	0.1979	0.0388	0.1932
Discontinuous Employment Dummy	0.0122	0.1099	0.0096	0.0976
0 Full Quarters Worked	0.0528	0.2237	0.0437	0.2043
1 Full Quarters Worked	0.0630	0.2429	0.0592	0.2360
2 Full Quarters Worked	0.0679	0.2516	0.0637	0.2443
3 Full Quarters Worked	0.0656	0.2476	0.0628	0.2425
4 Full Quarters Worked	0.7507	0.4326	0.7707	0.4204

Table 1.3: Sample Job Characteristics and Firm Interactions

	Men		Women	
		Standard		Standard
	Mean	Deviation	Mean	Deviation
Education Interacted with:				
ln(Firm Employment)	77.8903	35.5628	83.9931	34.4374
SIC 2-Digit Group (10-17)	0.9723	3.3710	0.2634	1.8376
SIC 2-Digit Group (20-29)	0.9511	3.3719	0.7315	2.9578
SIC 2-Digit Group (30-39)	2.2268	4.9013	1.1090	3.6003
SIC 2-Digit Group (40-49)	1.3239	3.9529	0.8601	3.2610
SIC 2-Digit Group (50-59)	2.6375	5.2537	2.1731	4.8323
SIC 2-Digit Group (60-67)	0.8346	3.3367	1.3848	4.0962
SIC 2-Digit Group (70-79)	1.4226	4.1722	1.1401	3.7530
SIC 2-Digit Group (80-89)	2.3280	5.2592	5.2352	6.7051
Non-White Interacted with:				
ln(Firm Employment)	1.7608	3.0250	2.1190	3.3776
SIC 2-Digit Group (10-17)	0.0176	0.1316	0.0057	0.0752
SIC 2-Digit Group (20-29)	0.0275	0.1635	0.0270	0.1620
SIC 2-Digit Group (30-39)	0.0552	0.2284	0.0352	0.1844
SIC 2-Digit Group (40-49)	0.0279	0.1647	0.0231	0.1502
SIC 2-Digit Group (50-59)	0.0639	0.2446	0.0511	0.2202
SIC 2-Digit Group (60-67)	0.0148	0.1206	0.0311	0.1736
SIC 2-Digit Group (70-79)	0.0376	0.1903	0.0319	0.1756
SIC 2-Digit Group (80-89)	0.0433	0.2036	0.1105	0.3135
Experience Interacted with:				
ln(Firm Employment)	135.7960	84.6476	143.5200	84.9780
SIC 2-Digit Group (10-17)	1.8232	6.7139	0.4793	3.5697
SIC 2-Digit Group (20-29)	1.7600	6.6453	1.3732	5.9306
SIC 2-Digit Group (30-39)	4.1983	9.9064	2.0685	7.1471
SIC 2-Digit Group (40-49)	2.4353	7.7308	1.4933	6.0308
SIC 2-Digit Group (50-59)	4.4610	9.6892	3.6959	9.0445
SIC 2-Digit Group (60-67)	1.2260	5.3552	2.1972	7.1314
SIC 2-Digit Group (70-79)	2.2161	7.1546	1.7777	6.4404
SIC 2-Digit Group (80-89)	3.8614	9.4503	8.8867	12.6636

firm-specific returns is diagonal. In Specification 3, the assumption of a diagonal variance matrix of firm-specific returns is relaxed in favor of an unstructured one that permits covariation between the firm-specific returns and intercepts. Finally, Specification 4 retains the assumption of an unstructured variance matrix of firm-specific components and introduces observed firm characteristics – firm size and industrial classification – into the characterization of γ_{jt}. Appendix C relates each of the specifications to the model developed in Section 1.3. Characteristics of the four specifications are summarized in Table 1.4.

The four specifications are chosen for a number of reasons. Specification 1 is common in the literature on the decomposition of wages into unobserved, but identifiable, firm- and person-effects. Specification 2 permits an evaluation of the importance of allowing for firm-specific deviations from the economy-wide returns to the elements of human capital and illustrates the relative importance of the estimated variance components. In Specification 3, the unstructured variance matrix of firm-specific returns permits an investigation of the significance of the covariation between the firm-specific returns and intercepts. Finally, Specification 4 introduces firm-level observed characteristics into the firm-specific returns in γ_{jt}. Estimation of this specification provides a direct assessment of the importance of both observed and unobserved firm characteristics in the determination of wages.

Specifications 1 through 3 are nested in terms of variance parameters and also share the same parameterization of the components for which coefficient estimates are fixed; for these specifications REML log-likelihood test statistics may be constructed to test which model best fits the data. Unfortunately, under the REML

Table 1.4: Model Descriptions

	Specification 1	Specification 2	Specification 3	**Specification 4**	Specification 5	Specification 6
Random Intercepts	YES	YES	YES	**YES**	YES	YES
Random Slopes	NO	YES	YES	**YES**	YES	NO
Variance Structure	DIAGONAL	DIAGONAL	UNSTRUCTURED	**UNSTRUCTURED**	DIAGONAL	DIAGONAL
Firm Observed Characteristics	NO	NO	NO	**YES**	YES	YES

approach, there is no corresponding test statistic for comparing models where the fixed component of the model changes. Thus, Specification 4, which introduces firm-level covariates into the wage model cannot be tested against the preceding specifications. However, Specification 4 may be tested against specifications with the more restrictive variance structures – like those in Specifications 2 and 1 – that also retain the firm observed characteristics. Table 1.4 describes two additional specifications that are estimated to test directly the superiority of Specification 4.

Specification 5 is identical to 4 except for the variance matrix of the firm error components which is now diagonal (as in Specification 2); Specification 6 is similar to 5, though, the firm-specific returns are removed and only the firm- and person-specific intercepts are retained. Estimates of the variance parameters and coefficients for Specifications 5 and 6 are not discussed in this paper, though, the following section provides statistics pertaining to model fit to motivate a preference for Specification 4.

For all specifications, observed determinants of wages are fully interacted with sex so that fixed coefficient estimates on all person, demographic, and firm characteristics are reported separately for men and women. The firm-specific random returns to the components of human capital are also fully interacted with sex; thus, for example, separate variances (and covariances) of the dispersion in the returns to education are reported for men and women. Finally, the random firm-intercepts are not assumed to differ across men and women.

1.5.2 Results

1.5.2.1 Model Selection

Hypothesis tests involving the variance components of competing models estimated using REML techniques are usually conducted by constructing REML likelihood ratio tests (REMLRT) which are only valid in cases where the parameterization of the fixed coefficients is the same in both models. In the current analysis, Specification 1 is nested, in terms of variance parameters, within Specification 2, while Specification 2 is nested within Specification 3. Specification 4, however, additional firm characteristics for which fixed coefficients are estimated. In this instance, an appropriate measure for model selection is the Bayesian Information Criterion (BIC).

Table **??** presents the Bayesian Information Criterion for each of the four specifications. Calculations of the BIC suggest that Specification 2 is preferred over Specification 1 and Specification 3 over Specification 2. Thus, an unstructured variance structure seems to fit the data best. Recall that Specification 4 uses a variance matrix of random coefficients identical to that of Specification 3 but also includes the firm covariates. The calculation of the BIC takes into account the inclusion of the additional fixed coefficients. Moreover, Specification 4 is preferred over Specification 5, which has the same variance parametrization as Specification 2; Specification 5 is preferred over Specification 6, which has the same variance parametrization as Specification 1. These results suggest that a fully unstructured variance matrix of firm returns is preferred, regardless of the choice of covariates to include in the fixed part of the model.

Table 1.5: Model Selection

	Specification 1	Specification 2	Specification 3	**Specification 4**	Specification 5	Specification 6
Random Intercepts	YES	YES	YES	**YES**	YES	YES
Random Slopes	NO	YES	YES	**YES**	YES	NO
Variance Structure	DIAGONAL	DIAGONAL	UNSTRUCTURED	**UNSTRUCTURED**	DIAGONAL	DIAGONAL
Firm Observed Characteristics	NO	NO	NO	**YES**	YES	YES
Log Likelihood	121133	124213	125747	**125832**	123531	117170
BIC	-242250	-248367	-251243	**-251413**	-247003	-234324
Test of Model Preference		2 over 1	3 over 2	**4 over 3 and 5**	5 over 6	

1.5.2.2 Coefficient and Variance Estimates

Recall that the estimation procedure first yields estimates of the model's variance parameters which are used to derive the fixed coefficient estimates – the BLUEs – and the random firm- and person-specific intercepts and firm-specific returns – the BLUPs. However, the estimates of the fixed coefficients are presented first as these are viewed, in the context of the firm-specific random returns, as the sample-wide average returns to the components of human capital. Thus, the discussion of the variance parameter estimates will be reviewed in light of these sample-wide averages. Predictors of the random components of the model are not reviewed in detail, though are used in the subsection that performs the analysis of variance decomposition.

Best Linear Unbiased Estimates (BLUEs)

Table 1.6 presents the fixed coefficient estimates (BLUEs) for Specifications 1, 2, and 3 as well as the Ordinary Least Squares (OLS) estimates of these coefficients. Recall that the data are demeaned prior to estimation, so no intercept (grand mean) is included in the estimations.

For Specification 1 (column 1), estimates of the fixed coefficients for men and women are reasonable in value and consistent with those found in the wage determination literature. For the quadratic in experience, the coefficient on the first order term is positive for both men and women, though, higher for men than for women; the coefficient on the second order term is negative for both men and women, though larger for women. The size of these estimates implies an experience profile for men that is higher and more steeply sloped than the one for women. The

Table 1.6: Fixed Coefficient Estimates for Specifications 1 through 3

	Specification 1		Specification 2		Specification 3		OLS	
Female								
Experience	0.0273	*	0.0277	*	0.0287	*	0.0224	*
	0.0010		0.0010		0.0014		0.0010	
$\frac{\text{Experience}^2}{100}$	-0.0424	*	-0.0427	*	-0.0455	*	-0.0397	*
	0.0019		0.0021		0.0029		0.0019	
Education	0.0370	*	0.0390	*	0.0430	*	0.0458	*
	0.0016		0.0017		0.0020		0.0009	
Non-White	-0.1584	*	-0.1695	*	-0.1688	*	-0.1700	*
	0.0087		0.0111		0.0111		0.0046	
Race Missing	-0.1229	*	-0.1321	*	-0.1404	*	-0.1834	*
	0.0224		0.0230		0.0230		0.0136	
Negative Experience	-0.0460		-0.0442		-0.0591		-0.0658	*
	0.0291		0.0295		0.0302		0.0198	
Male								
Constant	0.0490		0.0518		0.0574		0.1685	
	0.0336		0.0349		0.0371		33.9113	
Experience	0.0483	*	0.0488	*	0.0521	*	0.0408	*
	0.0009		0.0010		0.0014		0.0009	
$\frac{\text{Experience}^2}{100}$	-0.0764	*	-0.0764	*	-0.0826	*	-0.0645	*
	0.0018		0.0020		0.0030		0.0017	
Education	0.0420	*	0.0432	*	0.0458	*	0.0504	*
	0.0013		0.0014		0.0016		0.0007	
Non-White	-0.3307	*	-0.3111	*	-0.3230	*	-0.4173	*
	0.0083		0.0104		0.0106		0.0044	
Race Missing	-0.1857	*	-0.1682	*	-0.1607	*	-0.1827	*
	0.0184		0.0186		0.0188		0.0102	
Negative Experience	-0.0431		-0.0403		-0.0350		-0.1084	*
	0.0299		0.0304		0.0311		0.0201	
State								
State 1	-0.0886	*	-0.0779	*	-0.0799	*	-0.0921	
	0.0261		0.0264		0.0260		1.9011	
State 2	-0.0516	*	-0.0468	*	-0.0580	*	-0.1041	
	0.0175		0.0177		0.0175		1.9011	

* indicates significance at 5%

Non-White, Race Missing, and Negative Experience are dummy variables.

Standard errors appear below coefficient estimates.

return to education for men is positive and significant and implies a 4.2% increase in wages for each additional year of education; for women, the return to an additional year of education raises wages by 3.7%. The race dummy (non-white) is estimated to be negative for both men and women. Non-white men earn 33.07% less than white men; non-white women earn 15.84% less than white women.

The estimates for Specification 2, which includes the firm-specific returns to human capital, and for Specification 4, which includes the firm-specific returns to human capital as well as an unstructured variance matrix for these returns, are identical in sign and similar in magnitude. It is interesting, comparing the estimates from Specifications 1 through 3 to those emerging from OLS, that controlling for unobserved person- and firm-heterogeneity increases the wage penalty for non-white men; otherwise, coefficient estimates, where significant, are roughly the same magnitude.

Specification 4 includes the firm observed characteristics – firm employment and industry grouping – as well as their interactions with the observed components of human capital. Tables 1.7 and 1.7 present the estimates of these fixed coefficients for women and men, respectively. Rows identify the characteristics of workers in the wage model; columns identify firm characteristics. 2-digit SIC group (10-17), Mining and Construction, is the omitted industry group; thus, estimates under the "Constant" column heading refer to this group. Each row of estimates may be viewed as the effect of the observed firm characteristics on the wage impact of each observed component of human capital.

The first row suggests that for women, firm size has a significant effect on

Table 1.7: Fixed Coefficient Estimates for Specification 4: Women

		Firm Characteristics							
			2-Digit SIC Industry Group						
	Constant	ln(Employment)	(20-29)	(30-39)	(40-49)	(50-59)	(60-67)	(70-79)	(80-89)
Women									
Constant		0.0665 *	0.1275	0.0475	0.2598	-0.0393	-0.0381	-0.0448	-0.2573 *
		0.0130	0.1495	0.1384	0.1474	0.1299	0.1481	0.1373	0.1252
Experience	0.0251 *	0.0006	-0.0071	-0.0027	-0.0025	0.0001	-0.0024	0.0047	0.0014
	0.0082	0.0007	0.0087	0.0082	0.0090	0.0078	0.0084	0.0080	0.0076
$\frac{\text{Experience}^2}{100}$	-0.0408 *	-0.0009	0.0081	0.0037	-0.0020	-0.0038	0.0111	-0.0151	0.0002
	0.0169	0.0014	0.0177	0.0166	0.0184	0.0159	0.0173	0.0164	0.0154
Education	0.0431 *	-0.0008	-0.0032	-0.0043	-0.0163 *	-0.0020	0.0021	-0.0025	0.0155 *
	0.0083	0.0008	0.0083	0.0075	0.0079	0.0071	0.0086	0.0076	0.0068
Non-white	-0.1583 *	-0.0041	-0.0947	-0.0238	0.0662	0.0025	0.0669	-0.0293	0.0420
	0.0519	0.0048	0.0513	0.0476	0.0523	0.0457	0.0513	0.0467	0.0440
Race Missing	-0.1351 *								
	0.0230								
Negative	-0.0587								
Experience	0.0301								

* indicates significance at 5%

Non-White, Race Missing, and Negative Experience are dummy variables.

Standard errors appear below coefficient estimates.

average wages: the elasticity between firm size and real wages is 0.0665. Women in 2-digit SIC group (80-89) earn lower average wages than women working in other industries. The second and third rows suggest that the returns to experience are not significantly affected by firm size or industry group. The return to education for women is lower for women working in industry group (40-49), Transportation, Communications, etc., and higher for those working in industry group (80-89), Professional Services.

For men, the elasticity between firm employment and wages is significant and slightly lower than the estimate for women. And, the only significant industry grouping average wage effects are in: group (60-67), Finance, Insurance, and Real Estate; and group (70-79), Services; and group (80-89), Professional Services. Both the first order and second order components of the experience profile are significantly affected by firm size. Otherwise, the experience profile is only significantly different for men in Finance, Insurance, and Real Estate. The return to education is higher for men working in industry groups: (20-29), Manufacturing; (60-67) Finance, Insurance, and Real Estate; and (80-89) Professional Services. Finally, the non-white wage penalty is lower in Specification 4 than in Specifications 1 through 3.

One issue is whether the results from Specification 4 are driven by the mixed effects specification, specifically the inclusion of the random returns and intercepts, or by the inclusion of firm covariates and their interactions with worker characteristics. One way to check the robustness of the specification is to estimate the fixed coefficients of the model using OLS (assuming no random person or firm effects).

Table 1.8: Fixed Coefficient Estimates for Specification 4: Men and State Intercepts

		Firm Characteristics							
			2-Digit SIC Industry Group						
	Constant	ln(Employment)	(20-29)	(30-39)	(40-49)	(50-59)	(60-67)	(70-79)	(80-89)
Men									
Constant	0.2449	0.0548 *	-0.1107	-0.1232	0.1059	-0.1391	-0.4511 *	-0.2218 *	-0.3353 *
	0.1569	0.0114	0.0956	0.0863	0.0959	0.0800	0.1118	0.0854	0.0792
Experience	0.0329 *	0.0021 *	-0.0047	0.0022	-0.0039	0.0071	0.0170 *	0.0080	0.0094
	0.0053	0.0007	0.0056	0.0051	0.0060	0.0048	0.0064	0.0052	0.0050
$\frac{\text{Experience}^2}{100}$	-0.0501 *	-0.0040 *	0.0093	0.0005	0.0046	-0.0117	-0.0295 *	-0.0149	-0.0142
	0.0114	0.0015	0.0117	0.0107	0.0127	0.0102	0.0136	0.0110	0.0105
Education	0.0411 *	-0.0006	0.0111 *	0.0036	-0.0048	0.0016	0.0246 *	0.0073	0.0151 *
	0.0051	0.0007	0.0052	0.0047	0.0050	0.0044	0.0064	0.0046	0.0041
Non-white	-0.2413 *	-0.0073	-0.0625	-0.0349	-0.0221	-0.0534	-0.0827	-0.0592	-0.0280
	0.0354	0.0044	0.0355	0.0319	0.0365	0.0303	0.0425	0.0319	0.0306
Race Missing	-0.1585 *								
	0.0188								
Negative Experience	-0.0379								
	0.0311								
State									
State 1	-0.0609 *								
	0.0250								
State 2	-0.0738 *								
	0.0167								

* indicates significance at 5%

Non-White, Race Missing, and Negative Experience are dummy variables.

Standard errors appear below coefficient estimates.

In terms of the pricing of the components of human capital, equation (1.2) becomes

$$\gamma_{jt} = g'_{jt}\beta^{(2)} \tag{1.12}$$

and the fixed coefficient estimates retain the same meaning as in Specification 4.

Tables 1.9 and 1.10 present the estimates for the OLS estimation for women and men, respectively. In comparing the estimates in Tables 1.7 and 1.8 to those in Tables 1.9 and 1.10, notice that many of the values change in magnitude, in sign, and in level of significance suggesting that the results are sensitive to the estimation technique. For instance, the OLS estimate of the elasticity of log of real earnings (wages) and firm size is -0.0028 and insignificant compared to the significant estimate of 0.0665 in Specification 4.[16]

Overall, the fixed coefficient estimates emerging from Specifications 1 through 4 are consistent with the empirical wage determination literature. Men and women earn positive returns to education, have concave experience profiles, and earn a penalty for being non-white. Specification 4 suggests that large firms pay wage premia to both men and women. Finally, observed firm characteristics are shown to influence both average wages and, through their interaction with the observed components of human capital, are shown to influence a number of the returns to the observed components of human capital.

Variance Parameter Estimates

Tables 1.11 and 1.12 present estimates of the variance (and covariance) param-

[16]The sensitivity of these estimates will be examined in more detail in the next revision of this paper.

Table 1.9: OLS Coefficient Estimates Corresponding to Specification 4: Women

	Constant	Firm Characteristics: ln(Employment)	2-Digit SIC Industry Group: (20-29)	(30-39)	(40-49)	(50-59)	(60-67)	(70-79)	(80-89)
Women									
Constant		-0.0028	0.0705	0.0239	-0.0814	-0.5313 *	-0.2711 *	-0.1878	-0.5003 *
		0.0069	0.1253	0.1212	0.1254	0.1152	0.1197	0.1200	0.1124
Experience	0.0187 *	0.0023 *	-0.0245 *	-0.0169 *	-0.0050	-0.0079	-0.0065	-0.0154 *	-0.0198 *
	0.0071	0.0004	0.0075	0.0073	0.0076	0.0069	0.0072	0.0072	0.0068
$\frac{\text{Experience}^2}{100}$	-0.0521 *	-0.0027 *	0.0500 *	0.0356 *	0.0186	0.0179	0.0172	0.0230	0.0507 *
	0.0142	0.0008	0.0149	0.0146	0.0153	0.0139	0.0144	0.0145	0.0135
Education	0.0109	0.0003	0.0135 *	0.0182 *	0.0109	0.0261 *	0.0302 *	0.0219 *	0.0433 *
	0.0062	0.0004	0.0065	0.0063	0.0065	0.0060	0.0062	0.0063	0.0058
Non-white	-0.0937 *	0.0056 *	-0.3750 *	-0.2039 *	-0.0462	-0.1049 *	-0.1560 *	-0.2995 *	-0.0615
	0.0351	0.0020	0.0362	0.0350	0.0360	0.0338	0.0349	0.0351	0.0328
Race Missing	-0.1665 *								
	0.0132								
Negative Experience	-0.0821 *								
	0.0193								

* indicates significance at 5%

Non-White, Race Missing, and Negative Experience are dummy variables.

Standard errors appear below coefficient estimates.

Table 1.10: OLS Coefficient Estimates Corresponding to Specification 4: Men and State Intercepts

	Constant	Firm Characteristics							
			2-Digit SIC Industry Group						
		ln(Employment)	(20-29)	(30-39)	(40-49)	(50-59)	(60-67)	(70-79)	(80-89)
Men									
Constant	-0.1004	0.0599 *	-0.1347	-0.0943	-0.1220	-0.4894 *	-0.4601 *	-0.3441 *	-0.7161 *
	32.9043	0.0060	0.0681	0.0577	0.0636	0.0550	0.0709	0.0610	0.0581
Experience	0.0341 *	0.0014 *	-0.0115 *	-0.0143 *	-0.0066	0.0064	-0.0007	-0.0056	-0.0013
	0.0035	0.0004	0.0043	0.0037	0.0041	0.0035	0.0045	0.0038	0.0037
$\frac{\text{Experience}^2}{100}$	-0.0603 *	-0.0013	0.0168	0.0291 *	0.0174 *	-0.0103	-0.0060	-0.0028	-0.0016
	0.0070	0.0008	0.0085	0.0072	0.0081	0.0070	0.0092	0.0076	0.0074
Education	0.0555 *	-0.0046 *	0.0150 *	0.0200 *	0.0080 *	0.0198 *	0.0510 *	0.0187 *	0.0481 *
	0.0027	0.0003	0.0033	0.0029	0.0032	0.0027	0.0036	0.0031	0.0029
Non-white	-0.4588 *	0.0191 *	-0.0845 *	-0.0192	0.0626 *	-0.1012 *	-0.2137 *	-0.1945 *	-0.0182
	0.0186	0.0019	0.0216	0.0191	0.0214	0.0185	0.0244	0.0202	0.0196
Race Missing	-0.1713 *								
	0.0099								
Negative Experience	-0.1282 *								
	0.0195								
State									
State 1	-0.0472								
	1.8447								
State 2	-0.0981								
	1.8447								

* indicates significance at 5%

Non-White, Race Missing, and Negative Experience are dummy variables.

Standard errors appear below coefficient estimates.

Table 1.11: Variance Estimates: Firm and Person Intercepts and Residuals

	Firm Intercept ψ_1		Person Intercept α		Residual ϵ	
Specification (1)	0.2259	*	0.3228	*	0.0558	*
Specification (2)	0.2198	*	0.2986	*	0.0515	*
Specification (3)	0.2293	*	0.2897	*	0.0497	*
Specification (4)	0.2054	*	0.2897	*	0.0496	*

* indicates significance at 5%

eters for the Specifications 1 through 4. Table 1.11 provides estimates of the variance of the person-specific and firm-specific intercepts as well as the residuals; Table 1.12 presents the full set of firm-specific variance parameters for all specifications. For both Specifications 1 and 2, the variance matrices are diagonal and estimated values are reported in rows. For Specifications 3 and 4, the variance matrices are unstructured in terms of the firm-specific returns (and intercept). The diagonal elements refer to the estimated variance of each term. Estimated covariances are reported below the diagonal and correlations are reported above the diagonal.

Recall that Specification 1 that includes only firm- and person-specific intercepts is treated as the baseline model as it is common in the wage decomposition literature. Referring to Table 1.11, the estimates of the variance of the firm-specific intercept $\widehat{\sigma}^2_{\psi_1}$ (0.2259), the person-specific intercept $\widehat{\sigma}^2_{\alpha}$ (0.3228), and the residual $\widehat{\sigma}^2_{\varepsilon}$ (0.0558) are all statistically significant and have the following interpretation: that a one standard deviation increase in ψ_1, α, or ε increases wages by $\widehat{\sigma}_{\psi_1}$, $\widehat{\sigma}_{\alpha}$, or $\widehat{\sigma}_{\varepsilon}$ log points, respectively. The estimated variance of the person intercepts is larger than that of the firm intercepts, a common finding in the literature, suggesting that

Table 1.12: Variance Estimates: Firm Components

	Firm Intercept ψ_1	Male Non-white ψ_2	Male Education ψ_3	Male Experience ψ_4	Male $\frac{\text{Experience}^2}{100}$ ψ_5	Female Non-white ψ_6	Female Education ψ_7	Female Experience ψ_8	Female $\frac{\text{Experience}^2}{100}$ ψ_9
Specification (1)									
	0.2259 *								
Specification (2)									
	0.2198 *	0.0815 *	0.0004 *	0.0001 *	0.0010 *	0.0794 *	0.0003 *	0.0001 *	0.0005 *
Specification (3)									
ψ_1	**0.2293** *	-0.0606	0.1460	0.1934	-0.1112	-0.0840	0.1695	0.0970	-0.0792
ψ_2	-0.0092 *	**0.1005** *	-0.0787	0.0317	-0.0279	0.1815	0.0825	0.1356	-0.0960
ψ_3	0.0033 *	-0.0012	**0.0022** *	0.0068	0.1094	0.0435	0.5940	0.3054	-0.1862
ψ_4	0.0051 *	0.0006	0.0000	**0.0031** *	-0.9486	0.1171	0.4590	0.2189	-0.1084
ψ_5	-0.0062 *	-0.0010	0.0006 *	-0.0061 *	**0.0137** *	-0.1132	-0.3175	-0.1364	0.0681
ψ_6	-0.0126 *	0.0181 *	0.0006	0.0020 *	-0.0042 *	**0.0984** *	-0.0341	-0.0282	0.0672
ψ_7	0.0042 *	0.0014	0.0014 *	0.0013 *	-0.0019 *	-0.0006	**0.0027** *	-0.1079	0.2124
ψ_8	0.0021 *	0.0020 *	0.0007 *	0.0006 *	-0.0007 *	-0.0004	-0.0003 *	**0.0021** *	-0.9482
ψ_9	-0.0036 *	-0.0029	-0.0008 *	-0.0006 *	0.0008	0.0020	0.0010 *	-0.0041 *	**0.0090** *
Specification (4)									
ψ_1	**0.2054** *	-0.0852	0.1371	0.1817	-0.1118	-0.1062	0.1410	0.0955	-0.0982
ψ_2	-0.0123 *	**0.1013** *	-0.0749	0.0252	-0.0208	0.1781	0.0899	0.1348	-0.0989
ψ_3	0.0029 *	-0.0011	**0.0021** *	-0.0164	0.1268	0.0403	0.5787	0.3069	-0.1952
ψ_4	0.0045 *	0.0004	0.0000	**0.0030** *	-0.9493	0.1073	0.4454	0.2063	-0.1066
ψ_5	-0.0059 *	-0.0008	0.0007 *	-0.0060 *	**0.0135** *	-0.1018	-0.3073	-0.1237	0.0631
ψ_6	-0.0149 *	0.0175 *	0.0006	0.0018 *	-0.0037	**0.0957** *	-0.0379	-0.0329	0.0692
ψ_7	0.0033 *	0.0015	0.0014 *	0.0012 *	-0.0018 *	-0.0006	**0.0026** *	-0.1211	0.2118
ψ_8	0.0020 *	0.0020 *	0.0006 *	0.0005 *	-0.0007 *	-0.0005	-0.0003 *	**0.0021** *	-0.9503
ψ_9	-0.0042 *	-0.0030	-0.0009 *	-0.0005 *	0.0007	0.0020	0.0010 *	-0.0041 *	**0.0090** *

* indicates significance at 5%

For Specifications 1 and 2, the rows correspond to estimated variances of each parameter indicated in the column heading.

For Specifications 3 and 4, variance estimates are on the diagonal, covariance estimates are below the diagonal, and correlations are above the diagonal.

person-level heterogeneity is relatively more important than firm-level heterogeneity in wage variation.

The second, third, and fourth rows of Table 1.11 reveal only small changes in the estimated variances of the intercepts and residuals moving across specifications. With the exception of the slight increase in the estimated variance of the firm-intercepts going from Specification 2 to 3, the estimates decrease across specifications.

The second row in Table 1.12 reports parameter estimates for Specification 2 that includes the firm-specific returns to human capital and assumes a diagonal variance matrix for these pay policy parameters. All estimated parameters are statistically significant. The introduction of the firm-specific returns induces a slight decrease in the value of the estimated variance of the firm-specific intercepts.[17]

Also reported for Specification 2 are the estimated variances of the firm-specific returns to human capital. The estimated variance of the returns to race (non-white) is large for both non-white men and women. The interpretation of the estimates is similar to the interpretation of those for the firm-intercept. For example, for a non-white woman, a one standard deviation increase in the firm-specific return to being non-white ψ_6 increases wages by $\widehat{\sigma}_{\psi_6}$ log points. Moreover, the total variance in wages due to the variance of average wages for non-whites across firms is given by $\widehat{\sigma}^2_{\psi_1}+ \widehat{\sigma}^2_{\psi 2}$ for men and $\widehat{\sigma}^2_{\psi_1}+ \widehat{\sigma}^2_{\psi 6}$ for women. The measure of race used in this analysis is relatively crude, so the estimated dispersion of wages across firms

[17]Previous versions of this paper found a noticeable decrease in the estimated variance of the firm-intercepts in moving from Specification 1 to 2. The change in this result is due to the inclusion of state-specific intercepts and an intercept for male, which were omitted from the earlier analysis. The previously estimated variance components are included in Appendix D.1

for this characteristic should be interpreted with caution. Dispersion in the firm-specific returns to wages for non-whites may reflect labor market discrimination – in the sense that firms have flexibility in setting wages for non-whites – though would also arise if non-white workers sort into a broad range of both high- and low-wage firms or certain high- and low-wage occupations within firms. Recall that the estimated sample-wide return (the BLUE) to non-white for both men and women implies that an average wage penalty is associated with this characteristic; dispersion in the returns to being non-white simply captures the dispersion in this penalty across firms. Another way to think about dispersion in returns to non-white is to consider what it would mean if these variance parameter estimates were statistically insignificant (from zero). In this case, the return race could still be negative, though, would be identical across firms.

The estimated variance of returns to education and experience for Specification 2 suggest that the dispersion in the returns to education across firms for women is slightly higher than for men ($\widehat{\sigma}^2_{\psi_3} > \widehat{\sigma}^2_{\psi_7}$). Dispersion in the returns to education and experience is interpreted in a manner that is slightly different than for the firm- and person-specific intercepts and return to race. These parameters measure the dispersion in the firm-specific prices of the components of human capital, thus the change in wages attributed to a change in the return to education for men, for example, is given by $dw_{ijt} = d\psi_3 * (education)$ and depends on a worker's level of education. For a man with 16 years of education, a one standard deviation increase in the return to the firm-specific education increases his real wage by 0.32 log points.

The assumption of a diagonal variance matrix of firm-specific pay components is relaxed in Specification 3. Table 1.12 presents the firm-specific variance and covariance estimates for this fully unstructured variance matrix. Diagonal elements contain the estimated variances of the pay policy parameters. Estimated covariances are reported below the diagonal and correlations are reported above the diagonal. The first thing to notice is the change in the estimated variance components between Specifications 2 and 3. For both men and women, the estimated variances in the returns to race for both men and women increase as do the estimated variances of the return to education and the coefficients of the experience profiles. Thus, relaxing the structure of the variance matrix of pay policy parameters results in the attribution of more variation across firms in the returns to race, education, and experience. The estimation of covariance parameters permits a discussion of the extent to which firm pay policies are correlated within the sample. For instance, the first column of estimates for Specification 3 (below the first element of the diagonal) identify the covariation between the firm-specific intercepts and the firm-specific returns; the first row contains the correlations. The correlations between the firm-specific intercept and the returns to race for both men and women are -0.0606 and -0.0840, respectively. Firms that tend to pay high average wages to all employees relative to other firms also tend to pay lower wages to non-whites relative to other firms. For both men and women, high average firm wages are positively correlated with the returns to education. For both men and women, firm-specific intercepts are positively correlated with the return to the coefficient on the first order component of the experience profile and negatively correlated with the coefficient on the second order component suggesting that in high average wage firms, earnings profiles are relatively steeper for men for early years of experience but flatten in later years relative to the experience profiles of men at lower average wage firms. Other correlations reveal interesting characteristics of firm compensation policy. The correlation between men and women in the returns to race (0.1815), education (0.5940), and the terms of the experience profiles (0.2189 on the first order term and 0.0681 on the second) suggest that firms pursue somewhat similar pay policies across gender: firms that tend to reward highly the characteristics of women also reward highly these characteristics for men.

Finally, Table 1.12 reports the variance parameter estimates for Specification 4 which includes the firm observed characteristics. The terms on the diagonal change only slightly relative to those reported for Specification 3. This suggests that the inclusion of observed firm characteristics has little impact on the variance structure of the firm-specific returns.

The variance parameter estimates across the four specifications provide evidence that firms not only pay different wages to all of their workers but also pay different wages to certain types of workers. The estimated variances of the firm-specific returns to education, experience, and race are statistically significant and differ in size across men and women. The positive correlation across sex in the specific returns suggests a tendency for at least some firms to reward the human capital of both men and women similarly. Finally, firm average wages significantly co-vary with the returns to human capital. Non-whites tend to earn less at high average wage firms; education for both men and women are more highly rewarded in high average wage firms; and experience profiles become more steep, though exhibit turning points at lower levels of experience, in high average wage firms.

1.5.2.3 Correlations

Tables 1.13 and 1.14 provide the correlations between the dependent variable, the total person effect, the total firm effect, and the observed time-varying covariates. The total person effect is composed of two parts: (1) the estimated unobserved person-level intercept $\widehat{\alpha}_i$ and (2) an observed component capturing the average returns to education, race, the race missing dummy, and the negative experience dummy. The total firm effect is also composed of two parts: (1) the unobserved firm-level intercept $\widehat{\psi}_1$ and (2) the firm-specific returns to the components of human capital. In the first line, for Specification 1, both the total person effect and the unobserved component $\widehat{\alpha}_i$ are more highly correlated with the log of real earnings than are the firm-specific intercepts. Moreover, correlation between the log of real earnings and the observed time varying covariates (last column) is lower than the correlations between earnings and any other component. These are standard findings in the wage decomposition literature.

The introduction of the firm-specific random returns in Specification 2 changes slightly the correlations between the log of real earnings and the model components in Specification 1. The firm-specific returns to human capital, however, are positively correlated with the log of real earnings (0.28); this correlation is approximately the same size as the correlation between earnings and the observed component of the person effect as well as earnings and the returns to the time varying covariates

Table 1.13: Correlations for Specifications 1 and 2

	ln (Annualized Real Earnings	Total Person Effect	Unobserved Component (α)	Observed Component[a]	Total Firm Effect ($X^{(2)}\Psi$)	Firm Intercept (Ψ_1)	Firm-Specific Return[b]	Time-Varying Covariates ($X\beta$)
Specification 1								
ln(Annualized Real Earnings)	1.00	0.75	0.69	0.28	n.a	0.55	n.a	0.26
Total Person Effect		1.00	0.95	0.33		0.03		-0.06
Unobserved Component (α)			1.00	0.01		-0.01		-0.05
Observed Component				1.00		0.11		-0.04
Firm Intercept (ψ_1)						1.00		0.03
Time-Varying Covariates ($X\beta$)								1.00
Specification 2								
ln(Annualized Real Earnings)	1.00	0.71	0.66	0.27	0.62	0.56	0.28	0.27
Total Person Effect		1.00	0.94	0.35	0.06	0.05	0.02	-0.08
Unobserved Component (α)			1.00	0.01	0.02	0.02	0.02	-0.06
Observed Component				1.00	0.09	0.10	0.00	-0.07
Total Firm Effect ($X^{(2)}\Psi$)					1.00	0.90	0.44	0.03
Firm Intercept (ψ_1)						1.00	0.00	0.03
Firm-Specific Return							1.00	0.00
Time-Varying Covariates ($X\beta$)								1.00

[a] The observed component is the portion of wages attributed to: education, non-white (dummy), race missing (dummy), and negative experience (dummy).

[b] The firm-specific return is the portion of wages attributed to the firm-specific returns to person-level characteristics.

Table 1.14: Correlations for Specifications 3 and 4

	ln (Annualized Real Earnings	Total Person Effect	Unobserved Component (α)	Observed Component[a]	Total Firm Effect ($X^{(2)}\Psi$)	Firm Intercept (Ψ_1)	Firm-Specific Return[b]	Time-Varying Covariates ($X\beta$)
Specification 3								
ln(Annualized Real Earnings)	1.00	0.70	0.65	0.26	0.63	0.56	0.28	0.27
Total Person Effect		1.00	0.93	0.38	0.07	0.06	0.04	-0.09
Unobserved Component (α)			1.00	0.01	0.04	0.02	0.05	-0.05
Observed Component				1.00	0.08	0.11	-0.04	-0.09
Total Firm Effect ($X^{(2)}\Psi$)					1.00	0.89	0.44	0.01
Firm Intercept (ψ_1)						1.00	-0.02	0.03
Firm-Specific Return							1.00	-0.05
Time-Varying Covariates ($X\beta$)								1.00
Specification 4								
ln(Annualized Real Earnings)	1.00	0.68	0.65	0.22	0.59	0.52	0.28	0.34
Total Person Effect		1.00	0.94	0.34	0.05	0.04	0.04	-0.06
Unobserved Component (α)			1.00	0.01	0.04	0.01	0.05	-0.04
Observed Component				1.00	0.06	0.08	-0.03	-0.09
Total Firm Effect ($X^{(2)}\Psi$)					1.00	0.88	0.46	-0.03
Firm Intercept (ψ_1)						1.00	-0.02	-0.01
Firm-Specific Return							1.00	-0.04
Time-Varying Covariates (X β)								1.00

[a] The observed component is the portion of wages attributed to: education, non-white (dummy), race missing (dummy), and negative experience (dummy).

[b] The firm-specific return is the portion of wages attributed to the firm-specific returns to person-level characteristics.

$X\beta$. The firm-specific returns exhibit a correlation with the unobserved person component $\widehat{\alpha}$ that is nearly zero, suggesting that workers with high innate ability do not sort into firms that pay high specific returns.

Correlations for both Specifications 3 and 4, where the variance matrix of firm-specific returns is unstructured, change only slightly, and unremarkably, relative to Specification 2.

1.5.2.4 Analysis of Variance

The estimation approach – the use of REML to estimate the variance parameters and the Mixed Model Equations to predict and estimate the random and fixed coefficients – does not permit a straightforward decomposition of wages into variation attributable to the components of the wage model. For instance, there is no analog to a decomposition of variance using changes in model R-squared as in OLS. Even use of the estimated log likelihood function itself is of little value in assessing the relative importance of variation attributable to variables with estimated fixed coefficients and those with estimated random coefficients, as the value of the log likelihood itself is based only on the residual portion of the wage model. Nevertheless, this section attempts to assess the importance of the model's components in explaining variation in log wages through an analysis of variance exercise where the predicted random effects are treated as regressors in a linear regression on wages.

For each of the four specifications, a series of OLS regressions are estimated and include both the predicted random effects and the person-level covariates as

Table 1.15: Analysis of Variance

	Worker Covariates and Controls	Firm Covariates and Interactions	Predicted Person Intercept	Predicted Firm Intercept	Predicted Experience	Predicted Education	Predicted Non-White	Estimated Residual
Specification 1								
R-squared	0.1797	n.a.	0.6515	0.9205	n.a.	n.a.	n.a.	1.0000
Δ R-squared			0.4718	0.2690				0.0795
Specification 2								
R-squared	0.1797	n.a.	0.6028	0.8589	0.8970	0.9140	0.9279	1.0000
Δ R-squared			0.4231	0.2561	0.0381	0.0170	0.0139	0.0721
Specification 3								
R-squared	0.1797	n.a.	0.5915	0.8536	0.8817	0.9148	0.9310	1.0000
Δ R-squared			0.4118	0.2621	0.0281	0.0331	0.0162	0.0690
Specification 4								
R-squared	0.1797	0.2306	0.6395	0.8580	0.8840	0.9161	0.9319	1.0000
Δ R-squared		0.0509	0.4089	0.2185	0.0260	0.0321	0.0158	0.0681

regressors. Table 1.15 presents the results of this exercise.[18] For Specification 1, the baseline model $w_{ijt} = x_{ijt}^{(1)\prime}\beta^{(1)}$ is estimated yielding an R-squared of 0.1797. Introducing the predicted person intercept $\widehat{\alpha}_i$ increases the R-squared to 0.6515; adding the predicted firm intercept $\widehat{\psi}_1$ increases the R-squared to 0.9205; and, finally, including the estimated residuals $\widehat{\varepsilon}_{ijt}$ increases the R-squared to 1. Thus, observed worker characteristics (and control variables and time effects) explain roughly 18% of wage variation; predicted person intercepts explain roughly 47%; firm intercepts explain 27%; and the remaining 8% is explained by the residual.

For Specification 2, the firm-specific returns are included in the decomposition. The proportion of variation explained by the predicted person intercepts falls slightly to 42%; the proportion explained by the predicted firm intercepts is roughly 26%; and the proportion explained by the residual falls slightly to 7%. Collectively, the firm-specific returns to human capital account for around 7% of variation in wages: 3.8% is explained by the predicted returns to experience (for both men and women); 1.7% is explained by the predicted returns to education (for both men and women); and, 1.4% is explained by race (for both men and women). In Specification 3, the proportion of variation attributed to the firm-specific returns to human capital increases slightly due primarily to an increase in the proportion of variation attributed to education.

The observed firm characteristics are introduced in Specification 4 and explain a little over 5% of the variation in wages. The proportion of variation explained

[18] Results are relatively insensitive to the ordering of the components. Only one ordering is reported.

by the firm-specific returns to human capital is slightly over 7%; the proportion explained by firm intercepts alone is just over 21%. The relative size of these proportions underscores the importance of the firm-specific returns to human capital. The observed firm characteristics – firm size and industry division of operation – are broad measures in that they may not be precise enough to capture significant variation in wages across firms. For example, a larger sample of firms would permit the use of a finer measure of industry (*e.g.*, 4-digit SIC) which would likely capture more variation in earnings. That the proportion of variation attributed to the firm-specific returns is greater than the proportion accounted for by observed firm characteristics probably overstates the relative unimportance of firm observed characteristics. However, including additional and more precise firm characteristics would likely result in a reallocation of explanatory power from the firm-specific intercepts to the set of observed characteristics. Collectively, observed firm characteristics and firm-specific intercepts account for 26% of the variation in wages. Given the 7% of variation attributed to the firm-specific returns to human capital, the proportion of wages explained by the firm-specific returns is roughly 27% of the total proportion of wages *explained by firms*.

1.6 Conclusion

That firms pay different wages is a known characteristic of labor markets. The reasons for these differentials, however, are not as well understood. While the results in this paper support previous findings of the existence of firm average wage differentials, this papers shows that not only one type of firm-specific pay applies to all workers. Men and women, whites and non-whites, and experienced and inexperienced workers earn different wages at different firms.

The observed firm characteristics included in this analysis suggest that large firms pay higher wages and that industry, even broadly defined, captures variation in wages across firms. However, these observed firm characteristics explain only a small portion of wage variation across firms. Including more precise measures of firm-level characteristics – ones that would influence the wages of all workers at the firm – may help explain why firms pay high average wages to all workers, though, would not necessarily decrease the importance of why firms pay specific wages to specific workers.

Statistically significant variation in firm-specific pay across a variety of worker characteristics suggests that compensation policy is specific to both the firm and the workers it employs. Although firm-specific returns may reflect the sorting of workers into particular types of firms or occupations within firms, they may also reflect the way firms differentially value the human capital that these characteristics measure. Firms may adopt production technologies or strategies that require a particular mix of skills. The empirical literature on firm wage differentials suggests that firms that pay high average wages may also be more productive firms. If highly productive technologies require highly skilled workers, then it seems reasonable for a compensation policy to include high average wages and high returns to education. The correlation between these pay policy parameters would then reflect the extent to which these production technologies exist. Finally, if monitoring is difficult in certain types of firms, firms may pay higher than average wages to all workers. If certain types of workers, those who are highly educated and experienced, are more difficult to monitor or sort into jobs that are more difficult to monitor, then firms may pay high returns to these characteristics. The existence of variation in the returns to education and experience alone may reflect monitoring problems associated with these characteristics of workers. A positive correlation between these specific returns and firm average wages may suggest that monitoring is difficult in particular firms and doubly so for the highly educated and experienced workers employed by those firms.

Moving deeper into the firm by specifying a model that allows unobserved firm characteristics to influence both average wages as well as worker characteristics, I find that approximately 27% of the variation in wages attributed to firms arises from the specific returns that firms pay to the components of human capital. These returns are significantly dispersed across employers, exhibit strong correlations across firms, and appear to be more important than observed firm characteristics in explaining variation in wages. The underlying compensation strategies that this model captures, however, is still an open question. Tying the policy measures estimated in this paper to other firm-level outcomes, for example, turnover and productivity, may help answer this question.

Chapter 2

Confidentiality Protection in the Quarterly Workforce Indicators * †

2.1 Introduction

Disclosure proofing is the set of methods used by statistical agencies to protect the confidentiality of the identity of and information about the individuals and businesses that form the underlying data in the system. It is a critical component of any statistical system which uses confidential data to produce detailed public statistics, without compromising the confidentiality of the original micro-data.

Since 2003, the U.S. Census Bureau has published a new and novel statistical series: the Quarterly Workforce Indicators (QWI). The underlying data infrastructure was designed by the Longitudinal Employer-Household Dynamics Program at the Census Bureau (Abowd et al. 2004) and is described in detail elsewhere Abowd et al. (2005). At its core, the QWI system uses administrative records data col-

*This chapter is a reprint of: "Confidentiality Protection in the Quarterly Workforce Indicators," John M. Abowd, Bryce Stephens, and Lars Vilhuber. LEHD Technical Paper TP-2006-02

†This chapter reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureaus Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical, methodological, and technical issues are those of the author[s] and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov).

lected by a large number of states for both jobs and firms. This administrative database is enhanced with information from other micro-data sets at the Census Bureau. The QWI statistics offer unprecedented demographic and economic detail on the local dynamics of labor markets. Because of the fine detail offered by the published statistics and the confidential nature of the micro-data used to compile the statistics, confidentiality protection is a critical and integral part of the design of the QWI system. Only the application of state-of-the-art protection methods allows the Census Bureau to publish these statistics. In this article, we describe the fundamental components of the confidentiality protection system as it is used in the generation of the QWI. We also show that significant protection is provided by the system, but that the analytic validity of the data remains high. In particular, we provide evidence that the time-series properties of the disclosure-proofed data remain intact, and that the disclosure-proofed data is not biased.

In the QWI system, disclosure proofing is required to protect the information about individuals and businesses that contribute to the unemployment insurance (UI) wage records, the Quarterly Census of Employment and Wages (QCEW, also known as ES-202) reports,[1] and the Census Bureau demographic data that have been integrated with these sources. The primary concern of the confidentiality protection mechanism is thus with small cells, *i.e.*, cells that reflect data on few individuals or few firms.

Methods for protecting such data have been discussed before (*e.g.*, Cox and

[1] The Quarterly Census of Employment and Wages (QCEW) statistics are published by BLS in cooperation with state Labor Market Information offices.

Zayatz (1993)). In general, data are considered protected when "aggregate cell values do not closely approximate data for any one respondent in the cell" (Cox and Zayatz 1993, pg. 5). In the QWI confidentiality protection scheme, confidential micro-data are considered protected by noise infusion if one of the following conditions holds: (1) any inference regarding the magnitude of a particular respondent's data must differ from the confidential quantity by at least $c\%$ even if that inference is made by a coalition of respondents with exact knowledge of their own answers, or (2) any inference regarding the magnitude of an item is incorrect with probability no less than $y\%$, where c and y are confidential but generally "large." Condition (1) covers protection of magnitudes like total payroll. Condition (2) covers protection of counts assuming item suppression or some additional protection, like synthetic data, when the count is too small.

These two conditions are met by the multiple layers of confidentiality protection in the QWI system. The first layer occurs when job-level estimates are aggregated to the establishment level. A job-level measurement pertains to a given individual at a given workplace. As the job-level estimates are aggregated to the establishment level, the QWI system infuses specially constructed noise into the estimates of all of the workplace-level measures. This noise is designed to have three very important properties. First, every data item is distorted by some minimum amount. After this noise infusion, the distorted data item is used in all the publication QWIs. Second, for a given workplace, the data are always distorted in the same direction (increased or decreased) by the same percentage amount in every period. Third, the statistical properties of this distortion are such that when the estimates are aggregated, the effects of the distortion cancel out for the vast majority of the estimates, preserving both cross-sectional and time-series analytic validity. The use of multiplicative noise infusion, similar to what we develop here, as a cross-sectional confidentiality protection mechanism was first proposed by Evans et al. (1998).

A second layer of confidentiality protection occurs when the workplace-level measures are aggregated to higher levels, *e.g.,* sub-state geography and industry detail. The data from many individuals and establishment are combined into a (relatively) few estimates using a dynamic weight that controls the state-level beginning of quarter employment for all private employers to match the first month in quarter employment as tabulated from the Quarterly Census of Employment and Wages (QCEW). The establishment-level weight is used for every indicator in the QWIs. Hence, an additional difference between the confidential data item and the released data item arises from this weight The weighting procedure, combined with the noise infusion, move the published data away from the value contained in the underlying micro-data, and thus contribute to the protection of the confidentiality of the micro-data.

Third, some of the aggregate estimates turn out to be based on fewer than three persons or establishments. These estimates are suppressed and a flag set to indicate suppression. Suppression is only used when the combination of noise infusion and weighting may not distort the publication data with a high enough probability to meet the criteria laid out above. Count data such as employment are subject to suppression. Continuous dollar measures like payroll are not. All published estimates are still substantially influenced by the noise that was infused in the first layer of the protection system. These distorted estimates are published and flagged as substantially distorted. Each observation on any one of the published QWI tables thus has an associated flag that describes its disclosure status.

The remainder of this article is structured as follows. Section 2.2 describes the multiplicative noise model, and Section 2.3 details its integration into the computation of the QWI. The algorithm underlying the item suppression is outlined in Section 2.4, whereas the computation of weights is shown in Section 2.5. Sections 2.6 and 2.7 provide evidence on the extent of the protection and the analytic validity, respectively. Section 2.8 concludes.

2.2 Multiplicative noise model

To implement the multiplicative noise model, a random fuzz factor δ_j is drawn for each establishment j according to the following process:

$$p\left(\delta_j\right) = \begin{cases} \left(b-\delta\right)/\left(b-a\right)^2, \ \delta \in [a,b] \\ \left(b+\delta-2\right)/\left(b-a\right)^2, \ \delta \in [2-b, 2-a] \\ 0, \ \text{otherwise} \end{cases}$$

$$F(\delta_j) = \begin{cases} 0, \ \delta < 2-b \\ (\delta + b - 2)^2 / \left[2(b-a)^2\right], \ \delta \in [2-b, 2-a] \\ 0.5, \ \delta \in (2-a, a) \\ 0.5 + \left[(b-a)^2 - (b-\delta)^2\right], \ \delta \in [a, b] \\ 1, \ \delta > b \end{cases}$$

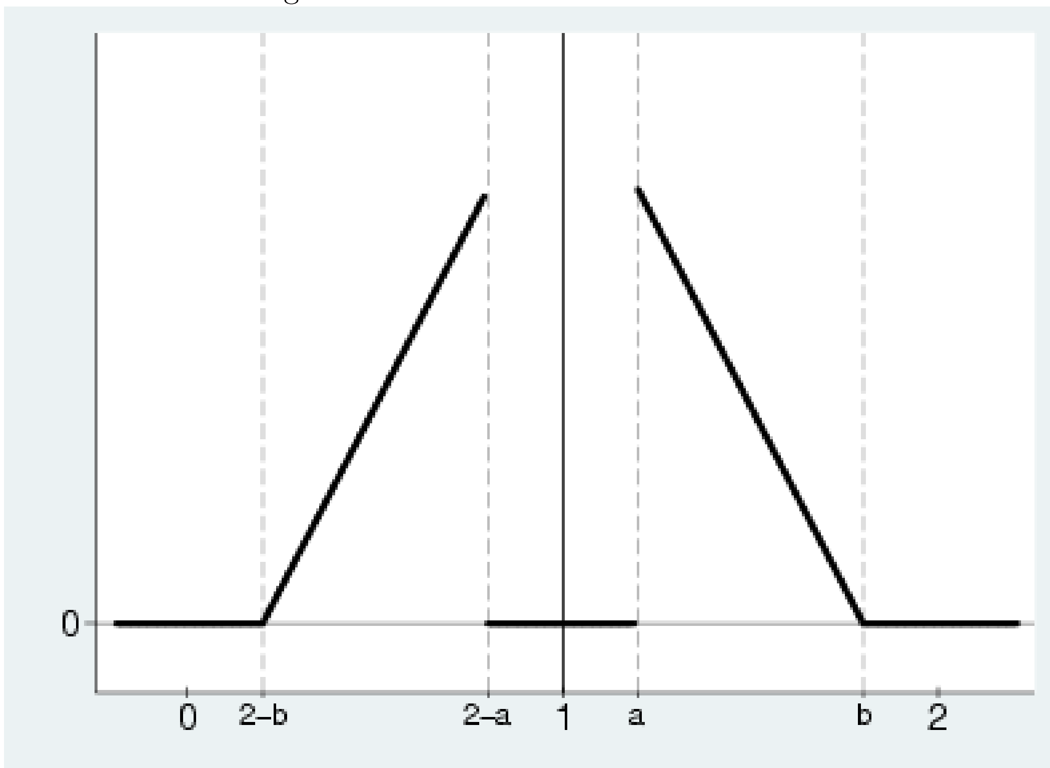
where $a = 1 + c/100$ and $b = 1 + d/100$ are constants chosen such that the true value is distorted by a minimum of c percent and a maximum of d percent.[2] Note that $1 < a < b < 2$. This produces a random noise factor centered around 1 with distortion of at least c and at most d percent. The distribution of δ is plotted in Figure 2.1.

A fuzz factor is drawn once for each employer, and for each of the establishments associated with that employer. Although fuzz factors vary across establishments, the fuzz factors attached to all establishments of the *same* employer are drawn from the same (upper or lower) tail of the fuzz factor distribution. Thus, if the fuzz factor associated with a particular employer is less than unity, then all of that employer's establishments will also have fuzz factors less than unity.

It is important to point out that fuzz factors are permanently attached to each employer and establishment and are retained for all time periods and for all revisions of QWI statistics.

[2] The exact numbers are confidential.

Figure 2.1: Distribution of Fuzz Factors



2.3 Applying the fuzz factors to estimates

Although all estimates are distorted based on the multiplicative noise model, the exact implementation depends on the type of estimate that is computed. A full discussion of how QWI estimates are computed can be found in Abowd et al. (2005), and a list of definitions for the statistics mentioned in this section, and the formulae for their computation is provided in Appendix E on page 121. In all cases, the noise infusion occurs at the level of an establishment estimate. By convention, distorted values are distinguished from their undistorted counterparts by an asterisk; *i.e.*, the true value of beginning-of-quarter employment is B and its distorted counterpart is B^*.

2.3.1 Distorting totals

The fuzz factor δ_j is used to distort all establishment totals by scaling of the true establishment level statistic

$$X^*_{jt} = \delta_j X_{jt},$$

where X_{jt} is an establishment level statistic among beginning-of-quarter (B), end-of-quarter (E) employment, flow employment (M), full-quarter employment (F), accessions (A), separations (S), new hires (H), recalls (R), flows into full-quarter status (FA), flows out of full-quarter status (FS), new hires into full-quarter status (FH), total payroll (W_1), payroll associated with E (W_2), with B (W_3), with new hires (WFH), periods of non-employment for accessions (NA), for new hires (NH), for recalls (NR), and for separations (NS).

2.3.2 Distorting averages of magnitude variables

Averages are constructed from distorted numerators (totals) with undistorted denominators according to

$$ZY^*_{jt} = \frac{Y^*_{jt}}{B(Y)_{jt}} = \delta_j \frac{Y_{jt}}{B(Y)_{jt}},$$

where ZY_{jt} is a statistic related to a total Y_{jt}, and $B(Y)$ is the appropriate denominator for the calculation of the average. Statistics distorted by this method are average earnings for various groups ($ZW_2, ZW_3, ZWFH, ZWA, ZWS$), and average periods of non-employment for several groups (ZNA, ZNH, ZNR, and ZNS).

2.3.3 Distorting differences of counts and magnitudes

Distorted net job flow (JF) is computed at the aggregate (k = geography, industry, or combination of the two for the appropriate age and sex categories) level as the product of the aggregated, undistorted rate of growth and the aggregated distorted employment:

$$JF^*_{kt} = G_{kt} \times \bar{E}^*_{kt} = JF_{kt} \times \frac{\bar{E}^*_{kt}}{\bar{E}_{kt}}.$$

This method of distorting net job flow will consistently estimate net job flow because it takes the product of two consistent estimators. The formulas for distorting gross job creation (JC) and job destruction (JD) are similar:

$$JC^*_{kt} = JCR_{kt} \times \bar{E}^*_{kt} = JC_{kt} \times \frac{\bar{E}^*_{kt}}{\bar{E}_{kt}}$$

and

$$JD^*_{kt} = JDR_{kt} \times \bar{E}^*_{kt} = JD_{kt} \times \frac{\bar{E}^*_{kt}}{\bar{E}_{kt}}.$$

where JCR_{kt} and JDR_{kt} are the aggregated growth rates for job creations and destructions, respectively. Exactly analogous expressions apply to full-quarter net job flows (FJF), full-quarter job creations (FJC), and full-quarter job destructions (FJD).

The same logic was used to distort wage changes for subgroups (accessions, separations, full-quarter accessions and separations). The undistorted total changes were divided by the undistorted denominators then multiplied by the ratio of the distorted denominator to the undistorted denominator for the computation of average change in earnings. Averages are distorted by multiplying by the ratio of the distorted denominator to the true denominator. For example:

$$Z\Delta WY_{kt}^{*} = \frac{\Delta WY_{kt}}{Y_{kt}} \times \frac{Y_{kt}^{*}}{Y_{kt}}.$$

where, again, Y denotes a particular count, and $Z\Delta WY$ the average change in total earnings associated with that particular count.

2.4 Item suppression

Despite the noise infusion described in the previous sections, some disclosure risk remains for counts based on very few entities in a cell. For counts based on data from fewer than three individuals or employers, the fuzz factors may not provide sufficient protection. This condition applies to the variables B, E, M, F, A, S, H, R, FA, FH, FS, JC, JD, JF, FJC, FJD, FJF. The QWIs therefore also implement item suppression based on the number of either workers or the number of employers that contribute data for that item in a cell k in time period t, where a cell represents a particular combination of geography × industry × age × sex. Because of the noise infusion used previously, however, no complementary suppressions are needed since all of the values based on three or more individuals or employers are adequately protected. Any estimate of the suppressed item computed by subtraction is also protected.

The algorithm for item suppression for these variables is as follows:

- Check the conditions leading to a disclosure flag of -2 or -1 (data availability). If met, set the item to missing in the release file.

- Determine whether the value can be computed according to Census standards:

  - For the variables JC, JD, and JF, (respectively, $FJC, FJD,$ and FJF) check whether the denominator average employment ($\bar{E}_{kt}$; respectively, $\bar{F}_{kt}$) in the relevant cell kt rounds to zero.
  - Check whether the item in cell kt rounds to zero.
  - Check whether the data used to construct the cell kt value were based on 1 or 2 individuals.
  - Check whether the data used to construct the cell kt value were based on 1 or 2 employers.

  If any of these conditions are met, set the disclosure status to 5 and set the item to missing in the release file.

- Check whether the distortion of cell kt value exceeds the limit set by the Census Disclosure Review Board[3]. If so, set the disclosure status to 9 and copy the distorted value to the release file.

- Otherwise, set the disclosure status to 1 and copy the distorted value to the release file.

[3]The precise value is confidential.

Table 2.1 lists all possible flag values.

Table 2.1: Disclosure flags in the QWI

Flag	Explanation
-2	no data available in this category for this quarter
-1	data not available to compute this estimate
0	no employment in this cell, or no positive denominator (OK to disclose a 0 for sum or count, missing for ratio)
1	OK, distorted value released
5	Value suppressed because it does not meet US Census Bureau publication standards.
9	data significantly distorted, distorted value released

2.5 Weighting the QWI

The economic concepts underlying the Quarterly Census of Employment and Wages (QCEW) statistics, published by BLS in cooperation with state Labor Market Information offices, and the QWI statistics, published by the U.S. Census Bureau, are similar, but not identical. While the QCEW reports employment on the 12th day of the month, for all months, as reported by employers for each establishment, the QWI has several measures of employment, all of which are derived from reports of quarterly employment and wages of individual workers at particular employers (state UI accounts). In particular, flow employment can be distinguished from point-in-time measures. Flow employment M_{jt} is defined as a simple count of employees who had positive, UI covered earnings and any time during quarter t at establishment j. Beginning of quarter employment B_{jt}, on the other hand, counts the number of employees present at establishment j in both quarter t and $t-1$, and by inference, on the 1st day of quarter t. By definition, flow employment will be higher than any point-in-time measure. The point-in-time measures in the QCEW and the QWI are comparable, and in particular, the QCEW report for employment on the 12th of the first month of a quarter (January, April, July, October) is comparable but not identical to the QWI measure of B.

These two measures are not identical because (a) they do not refer to exactly the same point in time, (b) the in-scope establishments differ slightly, and (c) they are computed from different universe data. The actual differences between these two measures are modeled and captured by the weighting scheme used in the QWI. To be precise, denote by $QCEW_{1,jt}$ the measured employment for the 12th of the first month on the QCEW report for establishment j in quarter t and let w_t denote the (state-specific) weight. Then the time-series of adjustment weights are defined by

$$w_t \sum_j b_{jt} = \sum_j QCEW_{1,jt} \tag{2.1}$$

for each time period t.

The weighting is not used to control to sub-state geography and industry because the characteristics of multi-unit employing establishments are multiply-imputed in the QWI data. Due to the way in which the UI wage records are collected at the state agencies, establishment identifiers are missing for multi-units. In the QCEW, sub-state geography and industry are coded directly at the establishment level.

The fact that workplace characteristics of geography and industry are multiply-imputed for multi-unit employers also has confidentiality protection implications. The establishment-level QWI micro-data for these entities were not provided by the responding firm (a UI account). Hence, there are no actual confidential micro-data measured at the establishment level. In effect, these establishments are protected by a form of synthetic data.

2.6 Extent of protection

The extent of the protection of the QWI micro-data can be measured by how many counts differ from their true values. The percentage deviation is a measure of the uncertainty about the true value that one can infer from the released value. The following tables show a series of comparisons designed to emphasize the contribution of each component of the QWI confidentiality protection mechanisms to the uncertainty about the true value. The contributions of weighting and noise-infusion can be separated by first comparing the undistorted, unweighted data with the undistorted, weighted data (Table 2.2), thus tabulating the number of cells that diverge from their true value solely due to weighting. The undistorted, weighted data are then compared to the distorted and weighted data (Table 2.3), highlighting the contribution of the noise infusion. Finally, a comparison of the undistorted, unweighted data with the published data (Table 2.4) brings out the combined contribution of weighting, noise infusion, and item suppression.

The tables display the row percentages and may be interpreted as the conditional probability of reporting the column entry given the row entry. A promi-

Table 2.2: Small Cells: B, Unweighted vs. Weighted

(a) Illinois

Unweighted count	*Weighted count* 0	1	2	3	4	5 or more
0	99.33	0.66	0.00	0.00	0.00	0.00
1	0.10	96.76	3.13	0.00	0.00	0.00
2	0.01	2.00	84.68	13.26	0.04	0.01
3	0.01	0.01	3.42	75.72	20.26	0.59
4	0.00	0.00	0.01	4.49	67.62	27.87
5 or more	0.00	0.00	0.00	0.01	0.59	99.39

Total number of cells: 14,229,968 . For details, see text.

(b) Maryland

Unweighted count	*Weighted count* 0	1	2	3	4	5 or more
0	99.10	0.90	0.00	0.00	0.00	0.00
1	0.11	94.36	5.52	0.00	0.00	0.00
2	0.04	0.53	73.83	25.45	0.13	0.02
3	0.03	0.03	1.42	55.47	41.79	1.25
4	0.02	0.02	0.04	1.85	41.39	56.69
5 or more	0.01	0.01	0.01	0.02	0.21	99.75

Total number of cells: 4,659,408 . For details, see text.

Table 2.3: Small Cells: B, Undistorted vs. Distorted

(a) Illinois

Undistorted count	*Distorted count* 0	1	2	3	4	5 or more
0	99.86	0.14	0.00	0.00	0.00	0.00
1	0.91	95.75	3.34	0.00	0.00	0.00
2	0.00	4.27	87.25	8.47	0.00	0.00
3	0.00	0.00	10.69	77.20	12.11	0.00
4	0.00	0.00	0.00	14.73	67.49	17.78
5 or more	0.00	0.00	0.00	0.00	1.93	98.07

Total number of cells: 14,229,968 . Both comparisons are for weighted data. For details, see text.

(b) Maryland

Weighted count	*Distorted count* 0	1	2	3	4	5 or more
0	99.83	0.17	0.00	0.00	0.00	0.00
1	0.73	92.35	6.91	0.00	0.00	0.00
2	0.00	5.07	80.45	14.48	0.00	0.00
3	0.00	0.00	12.51	71.21	16.27	0.00
4	0.00	0.00	0.00	17.62	65.74	16.63
5 or more	0.00	0.00	0.00	0.00	1.68	98.32

Total number of cells: 4,659,408 . For details, see text.

Table 2.4: Small Cells: B, Raw vs. Published

(a) Illinois

	Published count						
Unweighted count	Suppressed	0	1	2	3	4	5 or more
0	0.79	99.21	0.00	0.00	0.00	0.00	0.00
1	99.91	0.08	0.00	0.00	0.00	0.00	0.00
2	94.02	0.01	0.00	0.00	5.87	0.09	0.01
3	34.33	0.00	0.00	0.00	47.75	16.98	0.94
4	25.87	0.00	0.00	0.00	5.56	43.24	25.32
5 or more	15.20	0.00	0.00	0.00	0.03	0.82	83.95

Total number of cells: 14,229,968 . Raw is unweighted and undistorted. Published is after weighting, distorting, and suppression. For details, see text.

(b) Maryland

	Published count						
Unweighted count	Suppressed	0	1	2	3	4	5 or more
0	1.06	98.94	0.00	0.00	0.00	0.00	0.00
1	99.90	0.09	0.00	0.00	0.00	0.00	0.00
2	85.71	0.04	0.00	0.00	13.90	0.32	0.02
3	23.54	0.03	0.00	0.00	40.18	33.60	2.65
4	18.06	0.02	0.00	0.00	2.22	33.67	46.04
5 or more	8.44	0.01	0.00	0.00	0.02	0.26	91.26

Total number of cells: 4,659,408 . For details, see text.

nent feature of Tables 2.2 and 2.3 is the strong weight of the diagonal. The vast majority of cells is left unchanged by either noise infusion or weighting. Nevertheless, both weighting and noise infusion do affect significant number of cells. The changed cells in Table 2.2 are more likely to be found above the diagonal, demonstrating that the raw job-level wage records in the QWI system generally estimate lower beginning-of-quarter employment than month-one employment in the published establishment-record-based statistics in the QCEW. The changed cells in Table 2.3 are more symmetrically aligned around the diagonal, reflecting the symmetry of the noise distribution used to distort the data.

Table 2.4 shows the amount of suppression after weighting and noise-infusion as it relates to the original raw value. Note that all single-individual cells have been suppressed. This is not true for two-person cells, some of which have a weighted value that lies above the suppression threshold causing the weighted distorted estimate to be released. The converse is true for cells with three individuals. Due to weighting, some of these cells have weighted, undistorted values that lie below the suppression threshold, and are consequently suppressed. While not explicitly detailed in these tables, cells that contain count data based on fewer than three firms also generate suppressions, which are included in the suppression totals. Given the information in Tables 2.2 and 2.3, almost no cells with 4 or more individuals in the raw data have distorted and weighted data below 3 (a jump of two columns). Thus, for these cells, all suppressions are due to a small number of firms in a cell, or one of the other suppression conditions listed in Table 2.1. Overall, at the level of detail analyzed here (SIC3 $\times$ county $\times$ time $\times$ sex $\times$ age), around 25% of the beginning of period employment cells are suppressed in both the states analyzed here. For more aggregate tabulations, for instance at the SIC Division level, that percentage falls to between 5% and 10%.

Because total payroll, the other variable considered in detail in this paper, is a total (magnitude), not a count, it is never suppressed. The combination of weighting and distorting is sufficient to protect the confidentiality of this item without suppression because if the item is based on a single person or establishment, then the minimum distortion of the underlying micro-data applies. If the item is based on 2 employers or establishments then both micro-data items have been distorted at least the minimum percentage. Knowledge of one's own value does not help in inferring another's value because both data items were distorted in an unknown direction by an unknown minimum percentage. Even an accurate inference about one's own distortion factor supplies no information about the other parties' distortion factor, thus protecting that item by at least the minimum distortion factor in each direction.

2.7 Analytic validity

The noise infusion described in Section 2.2 is designed to preserve the analytic validity of the data. In order to demonstrate how successfully this validity has been preserved, we provide in this section evidence on the time-series properties of the distorted data, as well as evidence on the cross-sectional unbiasedness of the published data. In each case, we used data from Illinois and Maryland. We concentrate on two estimates, beginning-of-quarter employment B, and total payroll W_1. The unit of analysis is an interior sub-state geography × industry × age × sex cell kt. Sub-state geography in all cases is a county, whereas the industry classification is SIC. For our purposes, analytic validity is obtained when the data display no bias and the additional dispersion due to the confidentiality protection system can be quantified so that statistical inferences can be adjusted to accomodate it.

2.7.1 Time-series properties of distorted data

To analyze the impact on the time series properties of the distorted data, we estimated an AR(1) for the time series associated with each cell kt, using county-level data for all Illinois and Maryland counties. Two AR(1) coefficients are estimated for each cell-time series. The first order serial correlation coefficient computed using undistorted data is denoted by r. The estimate computed using the distorted data is denoted by r^*. For each cell, the error $\Delta r = r - r^*$ is computed. Table 2.5 on page 79 shows the distribution of the errors Δr across SIC-division × county cells, for B, A, S, F, and JF when comparing raw (confidential) data to distorted data, whereas Table 2.6 on page 80 compares the same variables between the raw and the published data, which excludes suppressed data items.

The tables show that the time series properties of all variables analyzed remain largely unaffected by the distortion. The maximum bias (as measured by the median of this distribution) is never greater than 0.001 (raw v. distorted or raw

v. published). The error distribution is tight: the semi-interquartile range of the distortion for B in Maryland is 0.010, which is less than the precision with which estimated serial correlation coefficients are normally displayed. The maximum semi-interquartile range for any variable in any one of the two states is 0.012[4].

The distribution of errors is similar when considering raw versus published data (Table 2.6). Furthermore, although the overall spread of the distribution is slightly higher when considering two-digit SIC $\times$ county and three-digit SIC $\times$ county cells, which are sparser than the SIC-division $\times$ county cells, the general results hold there as well. We conclude that the time series properties of the QWI data are unbiased with very little additional noise, which is, in general, economically meaningless.

Finally, correlation coefficients between the raw and published data are estimated for the time series associated with each cell kt. Table 2.9 provides the distribution of the correlation coefficients. The time series are highly correlated; the first percentile of the distribution of the correlation coefficients is no lower than 0.969 and the median is no lower than 0.990.

[4] The maximum semi-interquartile range for SIC2-based variables is 0.0241, and for SIC3-based variables, 0.0244.

Table 2.5: Distribution of the Error in the First Order Serial Correlation: SIC-division × County, Raw vs. Distorted Data

$$\Delta r = r - r^*$$

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
			IL SIC Division		
01	-0.069373	-0.049274	-0.052155	-0.066461	-0.007969
05	-0.041585	-0.031460	-0.032934	-0.039787	-0.004651
10	-0.028849	-0.022166	-0.023733	-0.027926	-0.002785
25	-0.011920	-0.009996	-0.010161	-0.011913	-0.001003
50	0.000571	0.000384	0.000768	0.000306	-0.000044
75	0.013974	0.011806	0.012891	0.012632	0.000776
90	0.030948	0.025152	0.026290	0.028299	0.002263
95	0.044233	0.033871	0.037198	0.040565	0.004375
99	0.078519	0.054415	0.060327	0.074212	0.007845
			MD SIC Division		
01	-0.059390	-0.050060	-0.049160	-0.048983	-0.010339
05	-0.032436	-0.030694	-0.030720	-0.028823	-0.004482
10	-0.022176	-0.023042	-0.023525	-0.018979	-0.002589
25	-0.009125	-0.010831	-0.010199	-0.007936	-0.001161
50	0.000658	0.000726	0.001123	0.000788	-0.000073
75	0.011639	0.012500	0.012871	0.010200	0.001044
90	0.024883	0.024917	0.024511	0.022358	0.002256
95	0.035014	0.033517	0.033028	0.030864	0.003699
99	0.059709	0.049903	0.050689	0.047204	0.008619

Unit of observation is a cell. Industry aggregation is SIC Division, geography aggregated to county level. For more details, see text.

Table 2.6: Distribution of the Error in the First Order Serial Correlation: County x SIC-division × County, Raw vs. Published Data

$$\Delta r = r - r^*$$

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
			IL County x SIC Division		
01	-0.085495	-0.092455	-0.098770	-0.079205	-0.008447
05	-0.047704	-0.046665	-0.045208	-0.046830	-0.004959
10	-0.034558	-0.031767	-0.032898	-0.033607	-0.003186
25	-0.015317	-0.014197	-0.015077	-0.015533	-0.001189
50	-0.000512	-0.000997	-0.000707	-0.001000	-0.000049
75	0.013438	0.011536	0.012457	0.011670	0.000861
90	0.030963	0.027037	0.028835	0.027970	0.002489
95	0.044796	0.037906	0.041862	0.040096	0.004801
99	0.080282	0.079122	0.083824	0.077419	0.007537
			MD County x SIC Division		
01	-0.065342	-0.072899	-0.072959	-0.058021	-0.009081
05	-0.035974	-0.036995	-0.040314	-0.030985	-0.004540
10	-0.024174	-0.027689	-0.028577	-0.021361	-0.002823
25	-0.010393	-0.013686	-0.012505	-0.009401	-0.001243
50	0.000230	-0.000542	0.000797	0.000279	-0.000025
75	0.011382	0.012628	0.013034	0.009429	0.001045
90	0.025160	0.026325	0.025272	0.022027	0.002799
95	0.035176	0.034114	0.034999	0.030152	0.004321
99	0.060042	0.056477	0.055043	0.049213	0.009208

Unit of observation is a cell. Industry aggregation is SIC Division, geography aggregated to county level. For more details, see text.

Table 2.7: Distribution of the Error in the First Order Serial Correlation: Two-digit SIC $\times$ County, Raw vs. Distorted Data

$$\Delta r = r - r^*$$

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
			IL SIC2		
01	-0.070671	-0.052107	-0.057965	-0.068505	-0.017139
05	-0.039739	-0.033252	-0.035271	-0.036607	-0.006337
10	-0.026348	-0.023354	-0.024951	-0.024729	-0.003599
25	-0.009891	-0.010622	-0.010718	-0.009530	-0.001238
50	0.000333	-0.000023	0.000675	0.000212	0.000003
75	0.012089	0.010960	0.013107	0.011015	0.001185
90	0.029082	0.025055	0.028222	0.026441	0.003455
95	0.042054	0.034896	0.038768	0.039589	0.005497
99	0.077996	0.058780	0.065105	0.072694	0.011871
			MD SIC2		
01	-0.056975	-0.055872	-0.057173	-0.049496	-0.014149
05	-0.033605	-0.035727	-0.037286	-0.029605	-0.006805
10	-0.023911	-0.025826	-0.027422	-0.020951	-0.003828
25	-0.009977	-0.011753	-0.012791	-0.008451	-0.001427
50	0.000075	0.000332	-0.000282	0.000140	0.000082
75	0.010242	0.012439	0.011353	0.008987	0.001532
90	0.024432	0.026786	0.025800	0.021818	0.004062
95	0.035468	0.035693	0.035284	0.031619	0.006035
99	0.061907	0.055054	0.055839	0.054744	0.011731

Unit of observation is a cell. Industry aggregation is SIC2, geography aggregated to county level. For more details, see text.

Table 2.8: Distribution of the Error in the First Order Serial Correlation: Two-digit SIC × County, Raw vs. Published Data

$$\Delta r = r - r^*$$

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
			IL SIC2		
01	-0.129094	-0.104500	-0.102003	-0.123819	-0.019439
05	-0.056734	-0.054465	-0.054423	-0.054914	-0.006630
10	-0.038474	-0.037901	-0.036443	-0.036726	-0.004058
25	-0.016431	-0.016847	-0.016628	-0.016082	-0.001277
50	-0.001610	-0.002131	-0.000789	-0.001742	0.000022
75	0.011486	0.011319	0.013833	0.010231	0.001235
90	0.029364	0.027751	0.031744	0.026192	0.003639
95	0.043912	0.039888	0.046670	0.040161	0.005915
99	0.082596	0.079321	0.098374	0.076498	0.014536
			MD SIC2		
01	-0.101585	-0.091941	-0.096422	-0.105893	-0.016338
05	-0.049849	-0.049707	-0.053894	-0.043979	-0.007201
10	-0.032742	-0.035509	-0.038168	-0.030164	-0.004159
25	-0.015218	-0.017011	-0.018759	-0.013736	-0.001780
50	-0.001978	-0.001817	-0.002780	-0.001532	0.000024
75	0.009548	0.013094	0.011995	0.008193	0.001590
90	0.024396	0.029727	0.028478	0.021555	0.004398
95	0.035172	0.041838	0.042422	0.032194	0.006325
99	0.065299	0.097201	0.105719	0.057076	0.012864

Unit of observation is a cell. Industry aggregation is SIC2, geography aggregated to county level. For more details, see text.

Table 2.9: Distribution of Correlation Coefficients: SIC Division × County, Raw vs. Published Data

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
			IL SIC DIVISION		
01	0.969	0.990	0.990	0.969	0.994
05	0.984	0.992	0.992	0.985	0.997
10	0.989	0.993	0.993	0.989	0.998
25	0.993	0.994	0.994	0.993	0.999
50	0.997	0.996	0.996	0.997	1.000
75	0.998	0.998	0.998	0.999	1.000
90	1.000	0.999	0.999	1.000	1.000
95	1.000	1.000	1.000	1.000	1.000
99	1.000	1.000	1.000	1.000	1.000
			MD SIC DIVISION		
01	0.982	0.991	0.991	0.980	0.995
05	0.989	0.993	0.993	0.989	0.998
10	0.992	0.994	0.994	0.992	0.999
25	0.995	0.995	0.995	0.995	1.000
50	0.998	0.997	0.997	0.997	1.000
75	0.999	0.998	0.998	0.999	1.000
90	1.000	0.999	0.999	1.000	1.000
95	1.000	1.000	1.000	1.000	1.000
99	1.000	1.000	1.000	1.000	1.000

Unit of observation is a cell. Industry aggregation is SIC Division, geography aggregated to county level.

2.7.2 Cross-sectional unbiasedness of the distorted data

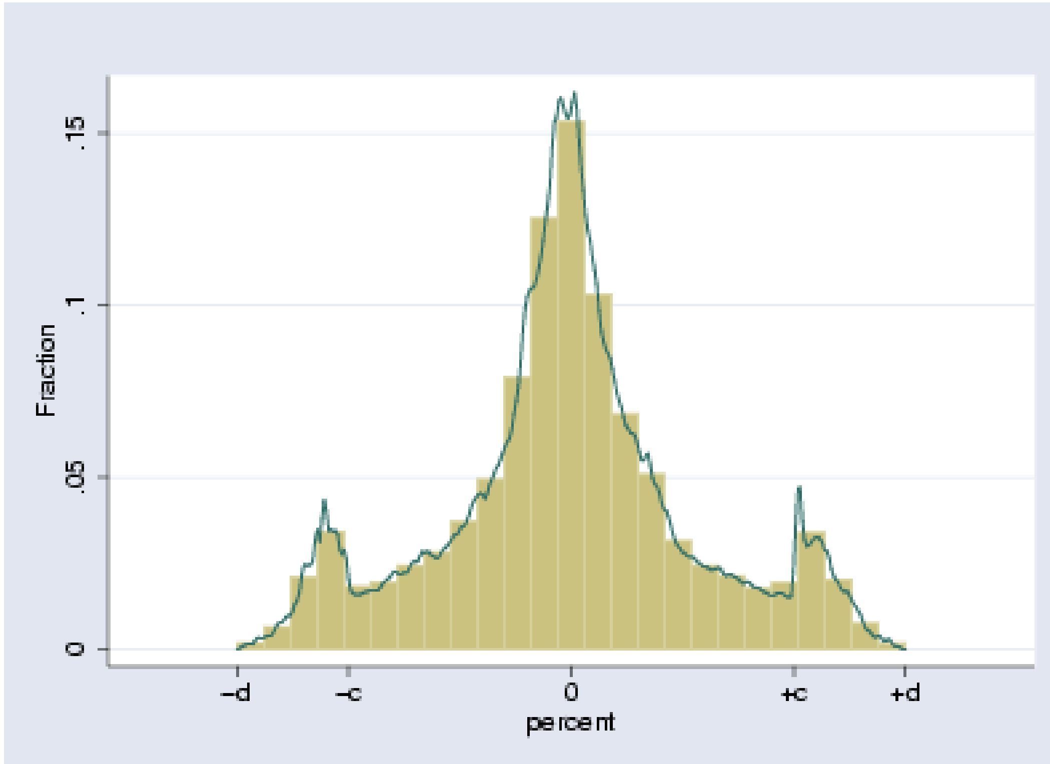
The distribution of the infused noise is symmetric, and allocation of the fuzz factors is random. The data distribution resulting from the noise infusion should thus be unbiased. Evidence of unbiasedness is provided by Figures 2.2 and 2.3. Each graph shows, for the states of Illinois (a) and Maryland (b) and a variable X, the distribution of the bias ΔX in each cell kt, expressed in percentage terms:

$$\Delta X_{kt} = \frac{X^*_{kt} - X_{kt}}{X_{kt}} \times 100 \qquad (2.2)$$

where X is B or W_1. All histograms are weighted by B_{kt}. Industry classification is three-digit SIC (industry groups).

Both the distribution of ΔB and ΔW_1 have most mass around the mode at zero percent. Also, as is to be expected, both present secondary spikes around $\pm c$, the inner bound of the noise distribution.

(a) Illinois



(b) Maryland

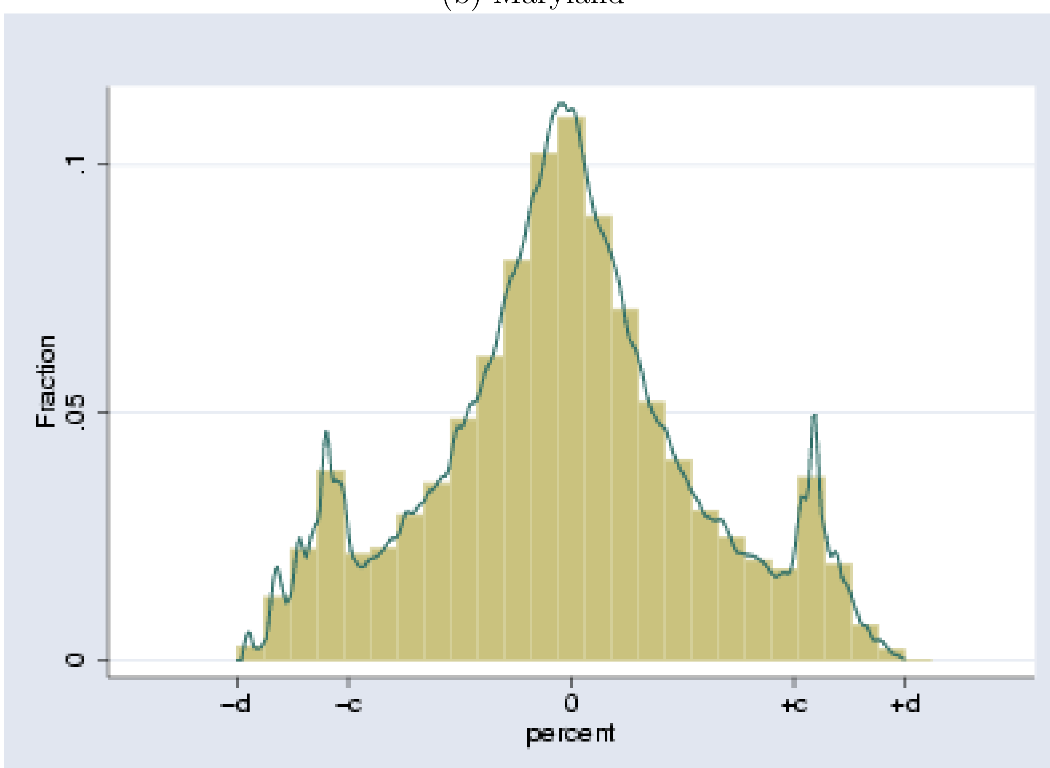
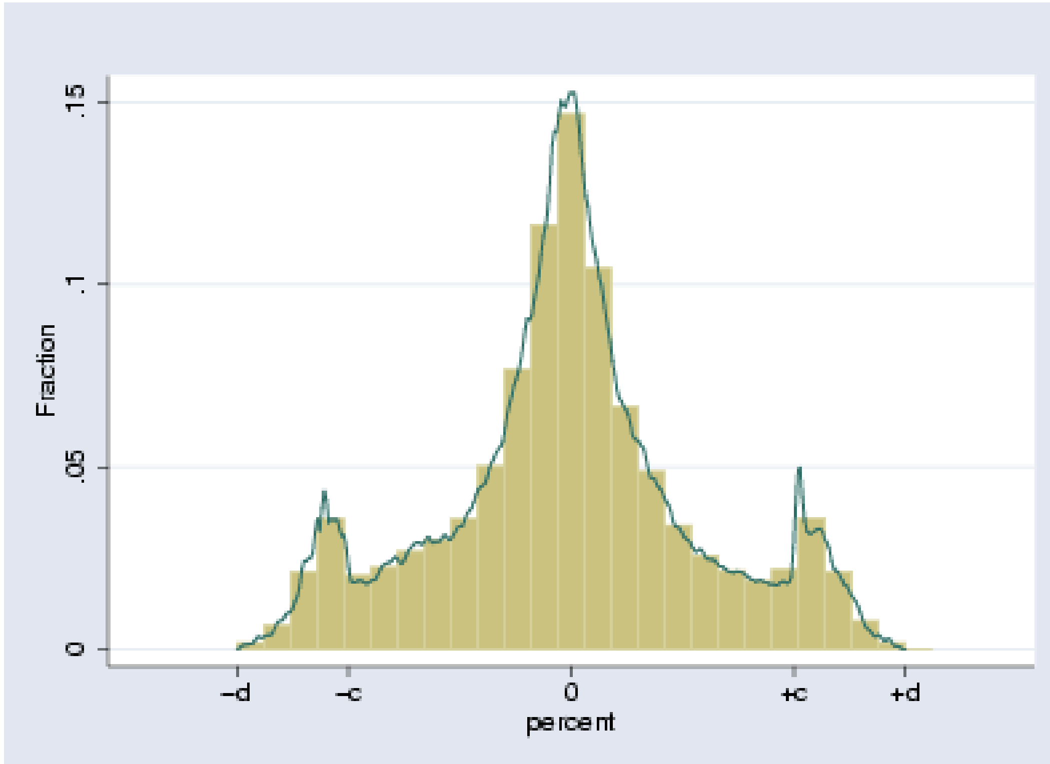


Figure 2.2: Distribution of Noise: B

(a) Illinois



(b) Maryland

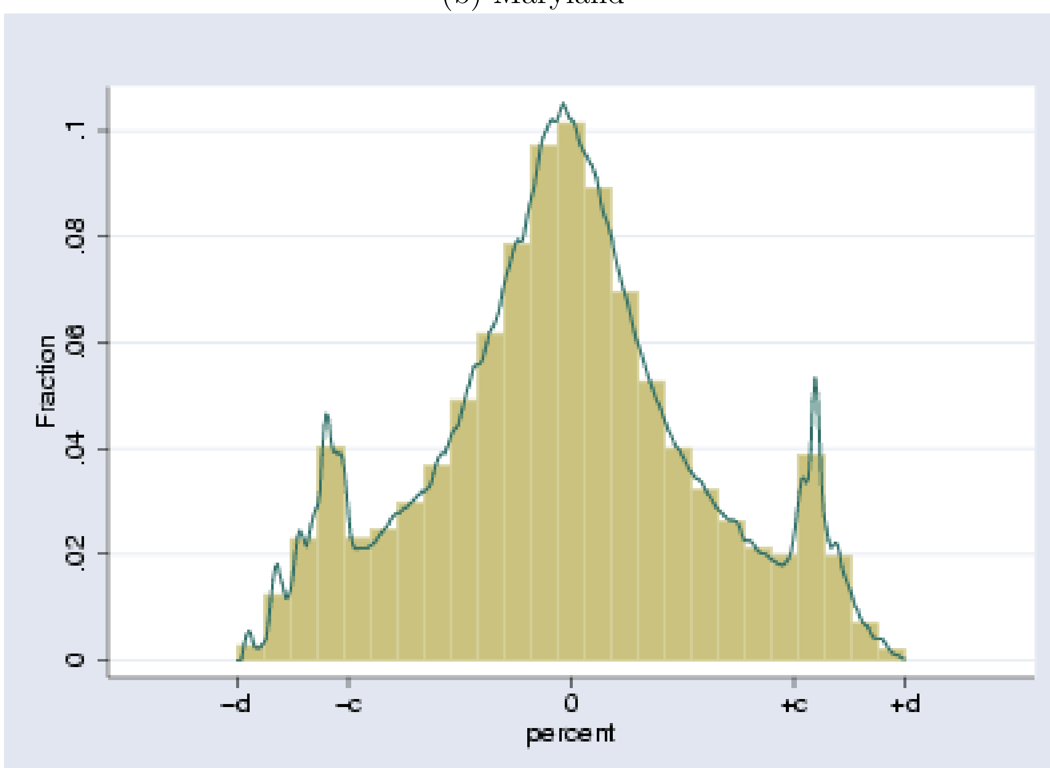


Figure 2.3: Distribution of Noise: W_1

2.8 Concluding remarks

In this paper, we provided a description of the confidentiality protection mechanism used in the generation of the Quarterly Workforce Indicators (QWIs). A notable feature of this disclosure proofing mechanism is the absence of table-level or complementary suppressions. Thus, although a significant number of count item values are indeed suppressed, the vast majority of counts are releasable data. All ratios and sums are released without suppression. To our knowledge, this is the first large-scale implementation of confidentiality protection by noise infusion.

Results from a comparison of the time-series characteristics of the undistorted and the distorted data shows remarkable consistency in the serial correlation coefficients between the two series at highly detailed levels. Furthermore, there is little or no bias in induced on average by the confidentiality protection mechanism, and the distributions of bias are tightly centered around the modal bias of zero.

Chapter 3

Imputation of Place of Work in the Quarterly Workforce Indicators: Implementation and Assessment * †

3.1 Introduction

A primary objective of the QWI is to provide employment, job and worker flow, and wage measures at a very fine level of geographic (place-of-work) and industry detail. The structure of the administrative data received by LEHD from state partners, however, poses a challenge to achieving this goal. QWI measures are primarily based on the processing of UI wage records which report, with the exception of Minnesota, only the employing firm of workers.[1] The QCEW micro-data, however, are comprised of establishment-level records which provide the level of ge-

*Sections 3.1 through 3.3 are taken directly from: "The LEHD Infrastructure Files and the creation of the Quarterly Workforce Indicators," John M. Abowd, Bryce Stephens, Lars Vilhuber with Frederik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. LEHD Technical Paper TP-2006-01.

†This chapter reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of research and to encourage discussion. This research is a part of the U.S. Census Bureaus Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01 AG018854, and the Alfred P. Sloan Foundation. The views expressed on statistical, methodological, and technical issues are those of the author[s] and not necessarily those of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers use of these data through the Research Data Centers (see www.ces.census.gov). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA. (Jeremy.S.Wu@census.gov http://lehd.dsd.census.gov).

[1]The employing firm is identified by a State Employer Identification Number (SEIN) that appears on individuals' wage records.

ographic and industry detail needed to produce the QWI. For firms operating only one establishment, the attachment of establishment-level characteristics is trivial. However, approximately 30 to 40 percent of state-level employment is concentrated in firms that operate more than one establishment. For these multi-establishment firms, an identifier on workers' wage records identifies the employing firm in the QCEW data, though, not the employing establishment.

In order to attach establishment-level characteristics to workers of multi-establishment firms, a probability model for employment location and imputation was developed. The model explains establishment-of-employment using two key characteristics available in the LEHD data: 1) distance between place-of-work and place-of-residence and 2) the distribution of employment across establishments of multi-establishment firms. The model is estimated using data from Minnesota, where both the firm and establishment identifiers appear on a worker's UI wage record. Then, parameters from this estimation are used to multiply impute establishment-of-employment for workers in the data from other states. Emerging from this process is an output file, called the unit-to-worker (U2W), containing ten imputed establishments for each worker of a multi-establishment firm. These implicates are then used in the downstream processing of the QWI. The statistical methods used to estimate the location model and to implement the multiple imputation are discussed in further detail in Rubin (1976), Rubin (1987), and Little and Rubin (1986).

The U2W process relies on information from five infrastructure files: (1) an Employer Characteristics File (ECF) containing information on firms and establishments, (2) a Geocoded Address List (GAL) file containing geography information for establishments and workers, (3) an Employment History File (EHF) containing individual earnings histories for workers in the UI universe, (4) an Individual Characteristics File (ICF) containing demographic information of workers in the UI universe, and (5) a Successor-Predecessor File (SPF) which provides a longitudinally consistent set of firm-level identifiers. Within the ECF, the universe of multi-establishment firms is identified. For these firms, the ECF also provides establishment-level employment, date-of-birth, and location (which is acquired from the GAL). The SPF contains information on predecessor relationships which may lead to the revision of date-of-birth implied by the ECF. Finally, individual work histories in the EHF in conjunction with place-of-residence information stored in the ICF provide the necessary worker information needed to estimate and apply the imputation model.

3.2 A Probability Model for Employment Location

3.2.1 Definitions

Let $i = 1, ..., I$ index workers, $j = 1, ..., J$ index firms (SEINs), and $t = 1, ..., T$ index time (quarters). Let R_{jt} denote the number of active establishments at firm j in quarter t, let $\mathfrak{R} = \max_{j,t} R_{jt}$, and $r = 1, ..., \mathfrak{R}$ index establishments. Note the index r is nested within j. Let N_{jrt} denote the quarter t employment of establishment r in firm j. Finally, if worker i was employed at firm j in t, denote by y_{ijt} the establishment at which she was employed.

Let $\mathcal{J}_t$ denote the set of firms active in quarter t, let $\mathcal{I}_{jt}$ denote the set of individuals employed at firm j in quarter t, let $\mathcal{R}_{jt}$ denote the set of active ($N_{jrt} > 0$) establishments at firm j in t, and let $\mathcal{R}^i_{jt} \subset \mathcal{R}_{jt}$ denote the set of active establishments that are feasible for worker i. Feasibility is defined as follows. An establishment $r \in \mathcal{R}^i_{jt}$ if $N_{jrs} > 0$ for every quarter s that i was employed at j.

3.2.2 The Probability Model

Let $p_{ijrt} = \Pr\left(y_{ijt} = r\right)$. At the core of the model is the probability statement:

$$p_{ijrt} = \frac{e^{\alpha_{jrt} + x'_{ijrt}\beta}}{\sum_{s \in \mathcal{R}^i_{jt}} e^{\alpha_{jst} + x'_{ijst}\beta}} \tag{3.1}$$

where α_{jrt} is a establishment- and quarter-specific effect, x_{ijrt} is a time-varying vector of characteristics of the worker and establishment, and β measures the effect of characteristics on the probability of being employed at a particular establishment. In the current implementation, x_{ijrt} is a linear spline in the (great-circle) distance between worker i's residence and the physical location of establishment r. The spline has knots at 25, 50, and 100 miles.

Using (3.1), the following likelihood is defined

$$p\left(y|\alpha, \beta, x\right) = \prod_{t=1}^{T} \prod_{j \in \mathcal{J}_t} \prod_{i \in \mathcal{I}_{jt}} \prod_{r \in \mathcal{R}^i_{jt}} \left(p_{ijrt}\right)^{d_{ijrt}} \tag{3.2}$$

where

$$d_{ijrt} = \begin{cases} 1 & \text{if } y_{ijt} = r \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

and where y is the appropriately-dimensioned vector of the outcome variables y_{ijt}, α is the appropriately-dimensioned vector of the α_{jrt}, and x is the appropriately-dimensioned matrix of characteristics x_{ijrt}. For α_{jrt}, a hierarchical Bayesian model based on employment counts N_{jrt} is specified.

The object of interest is the joint posterior distribution of α and β. A uniform prior on β, $p(\beta) \propto 1$ is assumed. The characterization of $p(\alpha, \beta|x, y, N)$ is based on the factorization

$$\begin{aligned} p(\alpha, \beta|x, y, N) &= p(\alpha|N)\, p(\beta|\alpha, x, y) \\ &\propto p(\alpha|N)\, p(\beta)\, p(y|\alpha, \beta, x) \\ &\propto p(\alpha|N)\, p(y|\alpha, \beta, x)\,. \end{aligned} \tag{3.4}$$

Thus the joint posterior (3.4) is completely characterized by the posterior of α and the likelihood of y in (3.2). Note (3.2) and (3.4) assume that the employment counts N affect employment location y only through the parameters α.

3.2.3 Estimation

The joint posterior $p(\alpha, \beta|x, y, N)$ is approximated at the posterior mode. In particular, we estimate the posterior mode of $p(\beta|\alpha, x, y)$ evaluated at the posterior mode of α. We compute the posterior modal values of the α_{jrt}, then, maximize the log posterior density

$$\log p\left(\beta|\alpha, x, y\right) \propto \sum_{t=1}^{T} \sum_{j \in \mathcal{J}_t} \sum_{i \in \mathcal{I}_{jt}} \sum_{r \in \mathcal{R}_{jt}^{i}} d_{ijrt} \left(\alpha_{jrt} + x_{ijrt}'\beta - \log\left(\sum_{s \in \mathcal{R}_{jt}^{i}} e^{\alpha_{jst} + x_{ijst}'\beta} \right) \right) \tag{3.5}$$

which is evaluated at the posterior modal values of the α_{jrt}, using a modified Newton-Raphson method. The mode-finding exercise is based on the gradient and Hessian of (3.5). In practice, (3.5) is estimated for three firm employment size classes: 1-100 employees, 101-500 employees, and greater than 500 employees, using data for Minnesota.

3.3 Imputing Place of Work

After estimating the probability model using Minnesota data, the estimated parameters are applied in the imputation process for other states. A description of the imputation method, as it relates to the probability model previously discussed, is provided in this section. Emphasis is placed on not only the imputation process itself but also the preparation of input data.

3.3.1 Sketch of Imputation Method

Ignoring temporal considerations, 10 implicates are generated as follows. First, using the mean and variance of β estimated from the Minnesota data, we take 10 draws of β from the normal approximation (at the mode) to $p\left(\beta|\alpha, x, y\right)$. Next, using QCEW employment counts for the establishments, we compute 10 values of

α_{jt} based on the hierarchical model for these parameters. Note these are draws from the exact posterior distribution of the α_{jrt}. The drawn values of α and β are used to draw 10 imputed values of place of work from the normal approximation to the posterior predictive distribution

$$p(\tilde{y}|x,y) = \int\int p(\tilde{y}|\alpha,\beta,x,y)\, p(\alpha|N)\, p(\beta|\alpha,x,y)\, d\alpha d\beta. \tag{3.6}$$

3.3.2 Implementation

3.3.2.1 Establishment Data

Using state-level micro-data, the set of firms (SEINs) that ever operate more that one establishment in a given quarter are identified; these SEINs represent the set of ever-multi-establishment firms defined above as the set $\mathcal{J}_t$. For each of these firms, its establishment-level records are identified. For each establishment, latitude and longitude coordinates, which emerge from GAL processing, parent firm (SEIN) employment, and QCEW first month employment[2] for the entire history of the establishment are retained. Those establishments with positive first-month employment in a given quarter characterize $\mathcal{R}_{jt}$, the set of all active establishments. An establishment date-of-birth is identified and, in most cases, is the first quarter in the QCEW time series in which the establishment has positive first-month employment. For some firms, predecessor relationships are identified in the SPF; in those instances, the establishment date-of-birth is adjusted to coincided with that

[2]In rare instances where no QCEW employment is available, an alternative employment measure based on UI wage record counts may be used.

of the predecessor's.

3.3.2.2 Worker Data

The EHF provides the earnings histories for employees of the ever-multi-establishment firms. For each in-scope job (a worker-firm pair), one observation is generated for the *end* of each job spell, where a job spell is defined as a continuum of quarters of positive earnings for worker at a particular firm during which there are no more than 3 consecutive periods of non-positive earnings.[3] The start-date of the job history is identified as the first quarter of positive earnings; the end-date is the last date of positive earnings.[4] These job spells characterize the set $\mathcal{I}_{jt}$.

3.3.2.3 Candidates

Once the universe of establishments and workers is identified, data are combined and a priori restrictions and feasibility assumptions are imposed. For each quarter of the date series, the history of every job spell that *ends in that quarter* is compared to the history of *every* active (in terms of QCEW first month employment) establishment of the employing firm (SEIN). The start date of the job spell is compared to the birth date of each establishment. Establishments that were born after the start of a job spell are immediately discarded from the set of candidate

[3]A new hire is defined in the QWI as a worker who acceeds to a firm in the current period but was not employed by the same firm in any of the 4 previous periods. A new job spell is created if, for example, a worker leaves a firm for 4 or more quarters and is subsequently re-employed by the same firm.

[4]By definition, an end-date for a job spell is not assigned in cases where a quarter of positive earnings at a firm is succeeded by fewer than 4 quarters of non-employment and subsequent re-employment by the same firm.

establishments. The remaining establishments constitute the set $\mathcal{R}^i_{jt} \subset \mathcal{R}_{jt}$ for a job spell (worker) at a given firm.[5]

Given the structure of the pairing of job spells with candidate establishments, it is clear that within job spell changes of establishment are ruled-out. An establishment is imputed once for each job spell thereby creating no false labor market transitions.[6]

3.3.2.4 Imputation and Output Data

Once the input data are organized, a set of 10 imputed establishment identifiers are generated for each job spell ending in every quarter for which both QCEW and UI wage records exist. For each quarter, implicate, and size class, $s = 1, 2, 3$, the parameters on the linear spline in distance between place-of-work and place-of-residence $\hat{\beta}^s$ are sampled from the normal approximation of the posterior predictive distribution of β^s conditional on Minnesota (MN)

$$p(\beta^s | \alpha_{MN}, x_{MN}, y_{MN}) \tag{3.7}$$

The draws from this distribution vary across implicates, but not across time, firms, and individuals.

[5]The sample of UI wage and QCEW data chosen for processing of the QWI is such that the start and end dates are the same. Birth and death dates of establishments are, more precisely, the dates associated with the beginning and ending of employment activity observed in the data. The same is true for the dates assigned to the job spells.

[6]More specifically, an establishment is imputed to a job spell only once within each implicate.

Next, for each firm j at time t, a set of $\hat{\alpha}_{jrt}$ are drawn from

$$p\left(\alpha_{ST}|N_{ST}\right) \tag{3.8}$$

which are based on the QCEW first-month employment totals (N_{jrt}) for all candidate establishments $r_{jt} \subset \mathcal{R}_{jt}$at firm j within the state (ST) being processed. The initial draws of $\hat{\alpha}_{jrt}$ from this distribution vary across time and firms but not across job spells.

Combining (3.7) and (3.8) yields

$$\begin{aligned} & p\left(\alpha_{ST}|N_{ST}\right) p(\beta^s|\alpha_{MN}, x_{MN}, y_{MN}) \\ \approx \; & p\left(\alpha_{ST}|N_{ST}\right) p(\beta^s|\alpha_{ST}, x_{ST}, y_{ST}) \\ = \; & p\left(\alpha_{ST}, \beta_{ST}|x_{ST}, y_{ST}, N_{ST}\right) \end{aligned} \tag{3.9}$$

an approximation of the joint posterior distribution of α and β^s (3.4) conditional on data from the state being processed.

The draws $\hat{\beta}^s$ and $\hat{\alpha}_{jrt}$ conjunction with the establishment, firm, and job spell data are used to construct the p_{ijrt} in (3.1) for all candidate establishments $r \in \mathcal{R}^i_{jt}$. For each job spell and candidate establishment combination, the $\hat{\beta}^s$ are applied to the calculated distance between place-of-residence (of the worker holding the job spell) and the location of the establishment, where the choice of $\hat{\beta}^s$depends on the size class of the establishment's parent firm. For each combination an $\hat{\alpha}_{jrt}$ is drawn which is based primarily on the size (in terms of employment) of the establishment relative to other active establishments at the parent firm. In conjunction, these determine the conditional probability p_{ijrt} of a candidate establishment's assignment to a given job spell. Finally, from this distribution of probabilities is drawn an establishment of employment.

Emerging from the imputation process is a data file containing a set of 10 imputed establishment identifiers for each job spell. In a minority of cases, the model fails to impute an establishment to a job spell. This is often due to unanticipated idiosyncrasies in the underlying administrative data. Furthermore, across states, the proportion of these failures relative to successful imputation is well under 0.5%. For these job spells, a dummy establishment identifier is assigned and in downstream processing, the employment-weighted modal firm-level characteristics are used.

3.4 Analysis

3.4.1 Analytic Validity

The file containing the 10 imputed establishment identifiers is used in conjunction with other worker and firm data to produce the 30 public use measures contained in the QWI for all state data in the LEHD Program except for Minnesota. For Minnesota, the recorded establishment of employment for multi-unit employers is used to create the QWI. One way of assessing the sensitivity of the QWI measures to the imputation of place of work is to compare the set of QWI measures for Minnesota that emerge from the regular processing of data using the recorded establishment of employment – call these the MNTRUE QWI – to a set of QWI measures generated where the employing establishment is treated as unknown – call these the MNIMPUTED QWI. This section provides an analysis of the discrepancy between the MNTRUE QWI versus the MNIMPUTED QWI for the universe of private employers for three levels of industry aggregation – SIC Division, 2-digit SIC, and 3-digit SIC – by county, age group, and sex for sets of QWI measures emerging from various points in the QWI production process.

The final set of 30 QWI measures that is published go through two stages of processing, after the imputation of establishment stage, that introduce variability in the measures. The first is the application of an establishment-level weight that is constructed to target state-level, quarterly Beginning of Quarter Employment to First Month Employment in the Quarterly Census of Employment and Wages published by the Bureau of Labor Statistics (BLS). This establishment-level weight is applied to the aggregation of all worker- and establishment-level QWI measures. The second is the application of additional confidentiality protections including: (1) the application of an establishment-level noise infusion factor (a "fuzz" factor) which distorts all measures tabulated using the raw and establishment-level weighted input data and (2) the suppression of data cells to which too few firms or employees contribute data. Moreover, the establishment-level fuzz factors are constructed so that for each state and quarter, the aggregate value of any QWI measure is approximately equal to the measure's undistorted value; however, at lower levels of aggregation (*e.g.*, a data cell for a specific sex and agegroup), it is more likely that the distorted values will deviate from the tabulated raw value.

The application of the establishment-level weights and fuzz factors to the QWI measures may introduce variability in the final QWI measures not due to imputation of establishment of work. Thus, discrepancies between the QWI measures for MNTRUE and MNIMPUTED that emerge from the final stage of processing are attributable to the imputation of establishment as well as the establishment-level weights and fuzz factors. In order to control for the impact of the establishment-level weights and fuzz factors on the bias statistics, as well as investigate the impact of the weights and factors on the discrepancies between the MNTRUE and MNIMPUTED QWI, three sets of QWI measures are generated for both MNTRUE and MNIMPUTED:

- Raw, No Establishment Weight Applied. This is a set of QWI measures generated without the application of the establishment-level weight and fuzz factor.

- Raw, Establishment Weight Applied. This is a set of QWI measures generated with the application of the establishment-level weight and without the establishment-level fuzz factor.

- Fuzzed, Establishment Weight Applied. This is a set of QWI measures generate with the application of the establishment-level weight and fuzz factor.

The 30 published QWI measures are listed in Table 3.1. Definitions and fundamental concepts related to the construction of the QWI measures are provided in Appendix E. For each interior data cell (defined by county, industry, agegroup, and sex) for each of the industry aggregations (SIC Division, 1-digit SIC, and 2-digit SIC) for all private employers over 1994Q3 through 2004Q3, a percentage or level bias statistic, measuring the discrepancy between the MNIMPUTED and MN-TRUTH is calculated. For the four core employment measures (B, E, F, and M), the level earnings measures, and the measures of periods of non-employment, the following percentage bias statistic is calculated:

$$\text{percentage bias} = \frac{\text{MNIMPUTED Value} \;-\; \text{MNTRUE Value}}{\text{MNTRUE Value}}$$

The measures for which the percentage bias is calculated, in the majority of data cells, take on non-zero values and, thus, a percentage bias measure is appropriate.

For the remaining measures, for which in a sizeable number of data cells a value of 0 is valid, a level bias measure is calculated:

$$\text{level bias} = \text{MNIMPUTED Value} \;-\; \text{MNTRUE Value}$$

Tables in Appendix F present statistics summarizing the distribution of the bias for the 30 QWI public use measures for SIC Division by 2-digit SIC and 3-digit SIC. In each table, there are 6 columns of data. The first two columns correspond to the "Raw, No Establishment Weight Applied" comparison of MNTRUTH to MNIMPUTED; the second two columns correspond to the "Raw, Establishment Weight Applied" comparison; and the third two columns correspond to the "Fuzzed, Establishment Weight Applied" comparison. In each set, the first column presents

Table 3.1: QWI Public Use Variables

QWI Variable	Variable Description	Bias Measure	Employment Variable Used in Bias Weighting
B	Beginning of Quarter Employment	Percent Deviation	B
E	End of Quarter Employment	Percent Deviation	E
F	Full Quarter Employment	Percent Deviation	F
M	Flow Employment	Percent Deviation	M
W1	Total Quarterly Earnings	Percent Deviation	M
Z_NA	Average Quarters of Non-Employment for Accessions	Percent Deviation	A
Z_NH	Average Quarters of Non-Employment for New Hires	Percent Deviation	H
Z_NR	Average Quarters of Non-Employment for Recalls	Percent Deviation	R
Z_NS	Average Quarters of Non-Employment for Separations	Percent Deviation	S
Z_W2	Earnings of End of Quarter Employees	Percent Deviation	E
Z_W3	Earnings of Full Quarter Employees	Percent Deviation	F
Z_WFA	Earnings of Full Quarter Accessions	Percent Deviation	FA
Z_WFS	Earnings of Full Quarter Separations	Percent Deviation	FS
Z_WH3	Earnings of Full Quarter Hires	Percent Deviation	H3
A	Accessions	Level Difference	n.a.
FA	Accessions to Full Quarter Status	Level Difference	n.a.
FJC	Full Quarter Job Creation	Level Difference	n.a.
FJD	Full Quarter Job Destruction	Level Difference	n.a.
FJF	Full Quarter Net Job Flow	Level Difference	n.a.
FS	Full Quarter Separations	Level Difference	n.a.
FT	Full Quarter Turnover	Level Difference	n.a.
H	New Hires	Level Difference	n.a.
H3	Full Quarter New Hires	Level Difference	n.a.
JC	Job Creation	Level Difference	n.a.
JD	Job Destruction	Level Difference	n.a.
JF	Net Job Flows	Level Difference	n.a.
R	Recalls	Level Difference	n.a.
S	Separations	Level Difference	n.a.
Z_dWA	Average Change in Total Earnings for Accessions	Level Difference	n.a.
Z_dWS	Average Change in Total Earnings for Separations	Level Difference	n.a.

summary statistics of the distribution of either the percentage or level bias; the second column provides weighted summary statistics for the percentage biases.[7]

Table F.1 presents summary statistics for the percentage bias of M, Flow Employment, for the SIC Division industry aggregation. M is simply a count of the number of wage records with positive earnings in a given quarter and is, in a sense, the purest employment measure created, as by definition it does not rely on the existence of previous or future quarters of positive earnings at a given establishment.[8] The first two columns present summary statistics on the percentage bias calculated using QWI data for which the establishment-level weights and fuzz factors are not applied. Notice that the mean of the percentage bias is close to zero and the median is equal to zero. The difference between the 90th and 10th percentile is 0.4040 and the inter-quartile spread is 0.1071. Weighting the percentage bias by M, results in a tightening of the distribution of the bias, reducing the difference between the 90th and 10th percentile to 0.2111 and the inter-quartile (75th minus 25th) spread to 0.0722. The tightening of the distribution due to weighting suggests that the imputation algorithm does a better job of allocating employment to larger data cells. The application of the establishment-level weights (the third and fourth data columns) and the establishment-level fuzz factors (the fifth and sixth columns) impact the distribution of the bias only slightly, leading to a small tightening of the difference between the 90th and 10th percentile as well as the inter-quartile spreads.

[7] Table 3.1 indicates which type of bias is calculated for each measure as well as the employment measure chosen in weighting the bias measures

[8] For instance, in order for a UI wage record to contribute to B, Beginning of Quarter Employment, there must be positive earnings a the employment establishment in both the previous and current quarter.

Table F.16 presents summary statistics for the percentage bias of M for the 2-digit SIC industry aggregation. Compared to the distribution of the bias for the SIC Division aggregation, the unweighted median remains at zero, though the unweighted mean increases to 1.35, most likely the influence of outliers accompanying the expansions of data cells in moving to a more disaggregated industry classification. Also, the inter-quartile spread is smaller and the difference between the 90th and 10th percentiles is larger moving from the SIC Division to 2-digit SIC industry aggregation suggesting a tightening of the center of the distribution and a widening in the tails. Weighting the proportional bias leads to a tightening of the difference between the 90th and 10th percentiles and a widening of the inter-quartile spread. Finally, the application of the establishment-level employment weight (columns 2 and 3) leads to a noticeable tightening of the distribution of the bias; this tightening is sustained when the fuzz factors are applied. The tightening of the distribution under the establishment-level weights suggests that the final QWI measures that are published are, in a sense, less biased than those that would emerge in the absence of the establishment-level weights.

Table F.31 presents summary statistics for the percentage bias of M for the 3-digit SIC industry aggregation. The dispersion in the bias increases as measured by an increase in the difference between the 90th and 10th percentile (1.5000) and the inter-quartile spread (0.2459). Again, this increase is most likely due to the further increase in data cells associated with moving to a lower level of industry aggregation. Finally, the introduction of the establishment-level weights leads to a noticeable tightening of the distribution of the bias, as before for the bias for the

2-digit SIC industry aggregation.

For the additional variables – B, E, F, W1, Z_NA, Z_NH, Z_NR, Z_NS, Z_W2, Z_W3, Z_WFA, and Z_WH3 – for which percentage bias measures are calculated, changes in the distribution of the bias across industry aggregations and due to the influence of the establishment-level weights and fuzz factors are similar to those for the distribution of the bias for M.

For the employment flow measures – A, FA, FJC, FJD, FJF, FS, H, H3, JC, JD, JF, R, S, and FT – a level bias is calculated. Changes in the distribution of the level bias for these variables are similar across industry aggregations. A discussion of the level bias for A, Accessions, illustrates. In Table F.8 for the SIC Division aggregate, the median value and the inter-quartile spread of the level bias is 0, reflecting, primarily, the prevalence of data cells – in both MNTRUE and MNIMPUTED – in which no accessions occur. The introduction of the establishment-level weights and fuzz factors induces a small widening on the left tail of the distribution. Moving across industry aggregates (Table F.23 and Table F.38), both the difference between the 90th and 10th percentile and the inter-quartile spread fall. This is due primarily to the increase in data cells in which no accessions occur.

Finally, for the two remaining measures of change in earnings – Z_dWA and Z_dWS in Tables F.15, F.30, and F.45– moving across industry aggregations leads to an increase in the absolute value of the average bias as well as a widening of the distribution of each bias. The introduction of the establishment-level weights leads to a tightening of the distribution of the bias across all industry aggregations.

One of the most interesting observations emerging from the comparison of the bias statistics across stages in QWI processing – particularly between the tabulations with and without the establishment-level weight applied – is the overall tightening of the distribution of the bias. The establishment-level weight is constructed to ensure that quarterly Beginning of Period Employment (B) is approximately equal to quarterly First Month Employment as published by the BLS. At the establishment-level, the weights, loosely speaking, are constructed so that establishment-level calculation of B is adjusted toward the recorded establishment-level First Month Employment (as recorded in the Quarterly Census of Employment and Wages micro-data). Thus, regardless of the nature of the input data to the QWI production process – *e.g.*, regardless of whether the production for Minnesota uses recorded or imputed establishment of work – the tabulated measure B is pushed toward the target value of First Month Employment at the establishment-level. Thus, the bias measures calculated for any data cell and for any QWI measures (assuming other measures are positively correlated with B) using the MNTRUE and MNIMPUTED QWI with the establishment weights applied should be less dispersed than the bias measures with the establishment weights not applied as measures are being pushed toward the same target value.

3.4.2 Time Series Correlation

Correlation coefficients between the raw and published data are estimated for the time series associated with each cell kt for B (Beginning-of-Period Employment), A (Accessions), S (Separations), F (Full Quarter Employment), and JF (Net Job

Flows). Table 3.2 provides the distribution of the correlation coefficients. The median of the distribution of the correlation coefficients is over 0.99 for all measures, with the exception of JF. The lower percentiles of the distribution suggest lower correlations for A, S, and JF relative to B and F.

Table 3.2: Distribution of Correlation Coefficients: SIC Division × County, Unweighted MNTRUE vs. Unweighted MNIMPUTED

Percentile	Beginning of Quarter Employment	Accessions	Separations	Full Quarter Employment	Net Job Flows
		Minnesota SIC Division			
01	0.547	0.198	0.155	0.583	0.112
05	0.582	0.366	0.469	0.619	0.266
10	0.637	0.653	0.689	0.783	0.503
25	0.988	0.968	0.979	0.986	0.767
50	0.996	0.990	0.990	0.996	0.903
75	0.999	0.994	0.995	0.998	0.976
90	0.999	0.998	0.998	0.999	0.989
95	1.000	0.998	0.998	1.000	0.993
99	1.000	0.999	0.999	1.000	0.998

Unit of observation is a data cell. Industry aggregation is SIC Division, geography aggregated to county level.

3.5 Conclusion

This paper provides a description of the development and implementation of the imputation of place-of-work in the Quarterly Workforce Indicators as well as an evaluation of the impact of the imputation process on the set of 30 public use measures published by the Longitudinal Employer-Household Dynamics Program, U.S. Bureau of the Census. The imputation process, which relies primarily on the distance between place-of-work and place-of-residence and the distribution of employment across establishments of multi-establishment firms to assign workers to place of work, is critical to the provision of the QWI measures at the finest possible level of demographic and geographic detail. This paper provides an description of the imputation process as well as an evaluation of the impact of the imputation process on the QWI public use measures. It is shown that the stage in QWI processing that introduces establishment-level weights in the tabulation of the public use measures also leads to a mitigation of the bias that arises from the imputation of place-of-work.

Appendix A

REML Estimation of Variance Parameters

A.1 Transforming the Data

REML estimation of the variance components requires performing maximum likelihood estimation on the portion of the dependent variable – wages – net of the model's fixed effects. The discussion that follows is based on the one provided in (Searle et al. (1992)).

First, multiply (1.8) by some vector k yielding:

$$k'w = k'X\beta + k'X^{(2)}\Psi + k'D\alpha \tag{A.1}$$

where:

$$k'X\beta = 0 \ \forall \ \beta \tag{A.2}$$

Thus:

$$k'X = 0 \tag{A.3}$$

The vector k' is of a specific form, described further in Appendix M.4e of (Searle et al. (1992)). The number of linearly independent vectors k' is determined by the order $(N \times (kl + m))$ and rank r of the matrix of covariates in X for which fixed effects are estimated. These vectors are collected in a matrix K so that $K'X = 0$.

A.2 Estimation of Variance Components

From equation (1.8) we have:

$$w \sim N(X\beta, V) \tag{A.4}$$

where:

$$V = X^{(2)}(I_J \otimes \Gamma)X^{(2)\prime} + \sigma_\alpha^2 DD' + \sigma_\varepsilon^2 I_N \tag{A.5}$$

Pre-multiplying by K' yields:

$$K'w \sim N(0, K'VK) \tag{A.6}$$

and substituting into the log likelihood:

$$\log L_R = \frac{1}{2}(N-r)\log 2\pi - \frac{1}{2}\log|K'VK| - \frac{1}{2}w'K\left(K'VK\right)^{-1}K'w \tag{A.7}$$

Differentiating the likelihood with respect to the variance parameters σ_α^2, σ_ε^2, and the elements of Γ yields the first order conditions (REML equations) for the maximization of the likelihood.[1]

[1] $\frac{\partial L_R}{\partial \sigma_\Gamma^2}$ refers generally to the derivative of the log likelihood with respect to each of the elements contained in Γ.

$$
\begin{aligned}
\frac{\partial L_R}{\partial \sigma_\alpha^2} &= -\frac{1}{2} trace \left[(K'VK)^{-1} K'DD'K \right] \\
&+\frac{1}{2} w'K (K'VK)^{-1} K'DD'K(K'VK)^{-1}K'w \\
\frac{\partial L_R}{\partial \sigma_\Gamma^2} &= -\frac{1}{2} trace \left[(K'VK)^{-1} K'X^{(2)}X^{(2)\prime}K \right] \\
&+\frac{1}{2} w'K (K'VK)^{-1} K'X^{(2)}X^{(2)\prime}K(K'VK)^{-1}K'w \\
\frac{\partial L_R}{\partial \sigma_\varepsilon^2} &= -\frac{1}{2} trace \left[(K'VK)^{-1} K'K \right] \\
&+\frac{1}{2} w'K (K'VK)^{-1} K'K(K'VK)^{-1}K'w
\end{aligned}
$$

Setting the first order conditions equal to zero and using $P = K(K'VK)^{-1}K'$ yields:

$$
\begin{aligned}
trace(PDD') &= w'PDD'Pw \qquad \text{(A.8)} \\
trace(PX^{(2)}X^{(2)\prime}) &= w'PX^{(2)}X^{(2)\prime}Pw \\
trace(P) &= w'PPw
\end{aligned}
$$

For the purpose of estimation, the conditions in (A.8) and the elements of the expected information matrix $\frac{\partial L_R}{\partial \sigma_i^2 \partial \sigma_j^2}$ for $i, j = \alpha, \Gamma, \varepsilon$ are evaluated using the Average Information (AI) algorithm discussed in Gilmore et al. (1995).

Appendix B

Derivation of Best Linear Unbiased Estimates (BLUEs) and Best Linear Unbiased Predictors (BLUPs)

Using the estimates of $\widehat{\sigma}^2_\alpha$, $\widehat{\sigma}^2_\varepsilon$, and the elements of $\widehat{\Gamma}$ that emerge from REML estimation, estimates of the fixed coefficients $\widehat{\beta}$ and the realized random coefficients $\widehat{\Psi}$ and $\widehat{\alpha}$ are obtained by maximizing the joint density of w, Ψ, and α with respect to β, Ψ, and α:

$$\max_{w.r.t.\ \beta,\Psi,\alpha} f(w, \Psi, \alpha)$$

and equating the first order conditions to zero, yielding:

$$\begin{bmatrix} X'\left[\frac{1}{\sigma^2_\varepsilon} I_N\right] X & X'\left[\frac{1}{\sigma^2_\varepsilon} I_N\right]\begin{bmatrix} X^{(2)} & D \end{bmatrix} \\ \begin{bmatrix} X^{(2)\prime} \\ D' \end{bmatrix}\left[\frac{1}{\sigma^2_\varepsilon} I_N\right] X & \begin{bmatrix} X^{(2)\prime} \\ D' \end{bmatrix}\left[\frac{1}{\sigma^2_\varepsilon} I_N\right]\begin{bmatrix} X^{(2)} & D \end{bmatrix} + \begin{bmatrix} I_J \otimes \Gamma & 0 \\ 0 & \sigma^2_\alpha I_I \end{bmatrix}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\Psi} \\ \widehat{\alpha} \end{bmatrix}$$

$$= \begin{bmatrix} X'\left[\frac{1}{\sigma^2_\varepsilon} I_N\right] w \\ \begin{bmatrix} X^{(2)\prime} \\ D' \end{bmatrix}\left[\frac{1}{\sigma^2_\varepsilon} I_N\right] w \end{bmatrix} \tag{B.1}$$

the Mixed Model (or Henderson) Equations. Solving yields the Best Linear Unbiased Estimates (BLUEs) $\widehat{\beta}$, the vector of fixed coefficients, and the Best Linear Unbiased Predictors (BLUPs) $\widehat{\Psi}$, the random firm coefficients and intercepts, and

$\widehat{\alpha}$, the random person intercepts:

$$\widehat{\beta} = BLUE\left[\beta\right] = \left(X'V^{-1}X\right)^{-1} X'V^{-1}w \tag{B.2}$$

$$\begin{bmatrix} \widehat{\Psi} \\ \\ \widehat{\alpha} \end{bmatrix} = BLUP \begin{bmatrix} \Psi \\ \\ \alpha \end{bmatrix} = \begin{bmatrix} \left(I_J \otimes \Gamma\right) X^{(2)\prime} & 0 \\ \\ 0 & \sigma_\alpha^2 I_I D' \end{bmatrix} V^{-1} \left(w - X\widehat{\beta}\right) \tag{B.3}$$

where:

$$V = \sigma_\alpha^2 DD' + X^{(2)} \left(I_J \otimes \widehat{\Gamma}\right) X^{(2)\prime} + \sigma_\varepsilon^2 I_N \tag{B.4}$$

In practice, estimates of the variance parameters σ_α^2, σ_ε^2, and Γ emerging from REML estimation are used to derive $\widehat{\beta}$, $\widehat{\Psi}$, and $\widehat{\alpha}$. Estimates of the residuals are given by:

$$\widehat{\varepsilon} = w - X\widehat{\beta} - X^{(2)}\widehat{\Psi} - D\widehat{\alpha} \tag{B.5}$$

Equation (B.1) also illustrates two key aspects of mixed model approach in the context of prediction in a model with random effects. First, as

$$\left| \begin{bmatrix} I_J \otimes \Gamma & 0 \\ \\ 0 & \sigma_\alpha^2 I_I \end{bmatrix}^{-1} \right| \rightarrow \infty \tag{B.6}$$

the solutions for $\widehat{\beta}$, $\widehat{\Psi}$, and $\widehat{\alpha}$ tend to the generalized least squares estimates where

$\widehat{\Psi}$ and $\widehat{\alpha}$ are treated as fixed instead of random. Given $\begin{bmatrix} I_J \otimes \Gamma & 0 \\ 0 & \sigma_\alpha^2 I_I \end{bmatrix}^{-1} < \infty$ the well known issues of identification that are present in models where $\widehat{\Psi}$ and $\widehat{\alpha}$ are treated as fixed are not present in the random effects case. Second, the off-diagonal elements of the left-hand-side of equation (B.1) reveal the impact of non-orthogonality assumption inherent in the modeling approach. These elements take into account the correlation of the designs of the person (D) and firm ($X^{(2)}$) effects with the covariates in X.

Appendix C

Details of Specifications

C.1 Specification 1

In Specification 1, the only firm-level random effect is an intercept. Thus, all covariates are included in $x_{ijt}^{(1)}$ for which parameters are fixed. $x_{it}^{(2)}$ contains only a constant.

$$x_{ijt}^{(1)} = \begin{bmatrix} \text{male*(race missing dummy)} \\ \text{male*(negative experience dummy)} \\ \text{(1-male)*(race missing dummy)} \\ \text{(1-male)*(negative experience dummy)} \\ \text{male*(experience)} \\ \text{male*(experience)\^{}2/100} \\ \text{male*(education)} \\ \text{male*(nonwhite dummy)} \\ \text{(1-male)*(experience)} \\ \text{(1-male)*(experience)\^{}2/100} \\ \text{(1-male)*(education)} \\ \text{(1-male)*(nonwhite dummy)} \\ \text{State 1} \\ \text{State 2} \\ \text{time effect 1} \\ \vdots \\ \text{time effect 43} \end{bmatrix}, x_{it}^{(2)} = \begin{bmatrix} \text{constant} \end{bmatrix} \tag{C.1}$$

In terms of the general model, conditions (1.4) become:

$$f_j = \begin{bmatrix} f_{1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ where } f_{1j} = 1, \Psi_j = \begin{bmatrix} \Psi_{1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Gamma = \begin{bmatrix} \sigma^2_{\Gamma_{11}} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \tag{C.2}$$

and equation (1.7) is estimated and, in terms of notation, simplifies to:

$$w_{ijt} = x_{ijt}^{(1)\prime}\beta^{(1)} + x_{it}^{(2)\prime}\beta^{(2)} + \Psi_j + \alpha_i + \varepsilon_{ijt} \tag{C.3}$$

where Ψ_j contains only one element, an intercept.

C.2 Specification 2

In Specification 2, the returns to the components of (and controls for) human capital are permitted to vary across firms and are contained in the vector $x_{it}^{(2)}$. Sample-wide average returns are estimated for the components in $x_{ijt}^{(1)}$ as well.

$$x_{ijt}^{(1)} = \begin{bmatrix} \text{male*(race missing dummy)} \\ \text{male*(negative experience dummy)} \\ \text{(1-male)*(race missing dummy)} \\ \text{(1-male)*(negative experience dummy)} \\ \text{State 1} \\ \text{State 2} \\ \text{time effect 1} \\ \vdots \\ \text{time effect 43} \end{bmatrix}, x_{it}^{(2)} = \begin{bmatrix} \text{constant} \\ \text{male*(nonwhite dummy)} \\ \text{male*(education)} \\ \text{male*(experience)} \\ \text{male*(experience)\^{}2/100} \\ \text{(1-male)*(nonwhite dummy)} \\ \text{(1-male)*(education)} \\ \text{(1-male)*(experience)} \\ \text{(1-male)*(experience)\^{}2/100} \end{bmatrix} \tag{C.4}$$

In terms of the general model, conditions (1.4) become:

$$f_j = \begin{bmatrix} f_{1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ where } f_{1j} = 1, \Psi_j = \begin{bmatrix} \Psi_{1j} \\ \vdots \\ \Psi_{lj} \end{bmatrix}, \Gamma = \begin{bmatrix} \sigma^2_{\Gamma_{11}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\Gamma_{ll}} \end{bmatrix} \tag{C.5}$$

where the vector Ψ_j captures the firm-specific returns to the elements of $x^{(2)}_{it}$. Equation (1.7) is estimated.

C.3 Specification 3

Specification 3 extends Specification 2 by permitting Γ in (C.5) to be unstructured.

$$f_j = \begin{bmatrix} f_{1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ where } f_{1j} = 1, \Psi_j = \begin{bmatrix} \Psi_{1j} \\ \vdots \\ \Psi_{lj} \end{bmatrix}, \Gamma = \begin{bmatrix} \sigma^2_{\Gamma_{11}} & \cdots & \sigma^2_{\Gamma_{1l}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{\Gamma_{l1}} & \cdots & \sigma^2_{\Gamma_{ll}} \end{bmatrix} \tag{C.6}$$

Equation (1.7) is estimated.

C.4 Specification 4

Finally, Specification 4 extends Specification 3 by modeling the firm-specific returns in equation (1.2) as a function of both the firm-specific random deviations in Ψ_j as well as observed firm-level employment and industry classification.

$$f_j = \begin{bmatrix} f_{1j} \\ \vdots \\ f_{lj} \end{bmatrix} \text{ where } f_{1j} = 1, \Psi_j = \begin{bmatrix} \Psi_{1j} \\ \vdots \\ \Psi_{lj} \end{bmatrix}, \Gamma = \begin{bmatrix} \sigma^2_{\Gamma_{11}} & \cdots & \sigma^2_{\Gamma_{1l}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{\Gamma_{l1}} & \cdots & \sigma^2_{\Gamma_{ll}} \end{bmatrix} \tag{C.7}$$

Equation (1.7) is estimated.

Appendix D

Variance Estimates: Previous Results

In an earlier version of this paper, an intercept for men and intercepts for employing U.S. state were not included in the fixed portion of the model. The most significant change between the estimates presented in this paper and those provided in earlier drafts is the magnitude of the change in the estimated variance of the firm-specific intercepts between Specifications 1 and 2. Table D.1 presents these estimates.

Notice that moving from Specification 1, which includes no firm-specific returns, to Specification 2, which does include firm-specific returns, the estimated variance of the firm-specific intercepts falls from 0.2606 to 0.1774 (and remains lower in Specifications 3 and 4). This decrease is not present in the current set of estimates presented in Chapter 1 in which the estimated models include an intercept for men and employing state.

Table D.1: Variance Estimates: Intercepts

	Firm Intercept ψ_1		Person Intercept α		Residual ϵ	
Specification (1)	0.2606	*	0.3215	*	0.0682	*
Specification (2)	0.1774	*	0.3080	*	0.0579	*
Specification (3)	0.1755	*	0.3022	*	0.0551	*
Specification (4)	0.1618	*	0.2946	*	0.0552	*

* indicates significance at 5%

Appendix E

Definitions of Fundamental LEHD Concepts

E.1 Dates

The QWI is a quarterly data system with calendar year timing. We use the notation YYYY:Q to refer to a year and quarter combination. For example, 1999:4 refers to the fourth quarter of 1999, which includes the months October, November, and December.

E.2 Employer

An employer in the QWI system consists of a single Unemployment Insurance (UI) account in a given state's UI wage reporting system. For statistical purposes the QWI system creates an employer identifier called an State Employer Identification Number (SEIN) from the UI-account number and information about the state (FIPS code). Thus, within the QWI system, the SEIN is a unique identifier within and across states but the entity to which it refers is a UI account.

E.3 Establishment

For a given employer in the QWI system, an SEIN, each physical location within the state is assigned a unit number, called the SEINUNIT. This SEINUNIT is based on the reporting unit in the ES-202 files supplied by the states. All QWI statistics are produced by aggregating statistics calculated at the establishment level. Single-unit SEINs are UI accounts associated with a single reporting unit in the state. Thus, single-unit SEINs have only one associated SEINUNIT in every quarter. Multi-unit SEINs have two or more SEINUNITS associated for some quarters. Since the UI wage records are not coded down to the SEINUNIT, SEINUNITs are multiply imputed as described in Abowd et al. (2005). A feature of this imputation system is that it does not permit SEINUNIT to SEINUNIT movements within the same SEIN. Thus, for multi-unit SEINs, the definitions below produce the same flow estimates at the SEIN level whether the definition is applied to the SEIN or the SEINUNIT.

E.4 Employee

Individual employees are identified by their Social Security Numbers (SSN) on the UI wage records that provide the input to the QWI. To protect the privacy of the SSN and the individual's name, a different branch of the Census Bureau removes the name and replaces the SSN with an internal Census identifier called a Protected Identity Key (PIK).

E.5 Job

The QWI system definition of a job is the association of an individual (PIK) with an establishment (SEINUNIT) in a given year and quarter. The QWI system stores the entire history of every job that an individual holds. Estimates are based on the definitions presented below, which formalize how the QWI system estimates the start of a job (accession), employment status (beginning- and end-of-quarter employment), continuous employment (full-quarter employment), the end of a job (separation), and average earnings for different groups.

E.6 Unemployment Insurance wage records (the QWI system universe)

The Quarterly Workforce Indicators are built upon concepts that begin with the report of an individual's UI-covered earnings by an employing entity (SEIN). An individual's UI wage record enters the QWI system if at least one employer reports earnings of at least one dollar for that individual (PIK) during the quarter. Thus, the job must produce at least one dollar of UI-covered earnings during a given quarter to count in the QWI system. The presence of this valid UI wage record in the QWI system triggers the beginning of calculations that estimate whether that individual was employed at the beginning of the quarter, at the end of the quarter, and continuously throughout the quarter. These designations are discussed below. Once these point-in-time employment measures have been estimated for the individual, further analysis of the individual's wage records results in estimates of full-quarter employment, accessions, separations (point-in-time and full-quarter), job creations and destructions, and a variety of full-quarter average earnings measures.

E.7 Employment at a point in time

Employment is estimated at two points in time during the quarter, corresponding to the first and last calendar days. An individual is defined as employed at the beginning of the quarter when that individual has valid UI wage records for the current quarter and the preceding quarter. Both records must apply to the same employer (SEIN). An individual is defined as employed at the end of the quarter when that individual has valid UI wage records for the current quarter and the subsequent quarter. Again, both records must show the same employer. The QWI system uses beginning and end of quarter employment as the basis for constructing worker and job flows. In addition, these measures are used to check the external consistency of the data, since a variety of employment estimates are available as point-in-time measures. Many federal statistics are based upon estimates of employment as of the 12th day of particular months. The Census Bureau uses March 12 as the reference date for employment measures contained in its Business Register and on the Economic Censuses and Surveys. The BLS Quarterly Census of Employment and Wages (QCEW)[1] series, which is based on the ES-202 data, use the 12th of each month as the reference date for employment. The QWI system cannot use exactly the same reference date as these other systems because UI wage reports do not specify additional detail regarding the timing of the wage payments. QWI research has shown that the point-in-time definitions used to estimate beginning and end of quarter employment track the QCEW month one employment estimates

[1] The QCEW were formerly known as Covered Employment and Wages (CEW).

well at the level of an employer (SEIN). For single-unit SEINs, there is no difference between an employer-based definition and an establishment-based definition of point-in-time employment. For multi-unit SEINs, the unit-to-worker imputation model assumes that unit-to-unit transitions within the same SEIN cannot occur. So, point in time employment defined at either the SEIN or SEINUNIT level produces the same result.

E.8 Employment for a full quarter

The concept of full quarter employment estimates individuals who are likely to have been continuously employed throughout the quarter at a given employer. An individual is defined as full-quarter-employed if that individual has valid UI-wage records in the current quarter, the preceding quarter, and the subsequent quarter at the same employer (SEIN). That is, in terms of the point-in-time definitions, if the individual is employed at the same employer at both the beginning and end of the quarter, then the individual is considered full-quarter employed in the QWI system.

Consider the following example. Suppose that an individual has valid UI wage records at employer A in 1999:2, 1999:3, and 1999:4. This individual does not have a valid UI wage record at employer A in 1999:1 or 2000:1. Then, according to the definitions above, the individual is employed at the end of 1999:2, the beginning and end of 1999:3, and the beginning of 1999:4 at employer A. The QWI system treats this individual as a full-quarter employee in 1999:3 but not in 1999:2 or 1999:4. Full-quarter status is not defined for either the first or last quarter of available data.

E.9 Point-in-time estimates of accession and separation

An accession occurs in the QWI system when it encounters the first valid UI wage record for a job (an individual (PIK)-employer (SEINUNIT) pair). Accessions are not defined for the first quarter of available data from a given state. The QWI definition of an accession can be interpreted as an estimate of the number of new employees added to the payroll of the establishment (SEINUNIT) during the quarter. The individuals who acceded to a particular employer were not employed by that employer during the previous quarter but received at least one dollar of UI-covered earnings during the quarter of accession.

A separation occurs in the current quarter of the QWI system when it encounters no valid UI wage record for an individual-employer pair in the subsequent quarter. This definition of separation can be interpreted as an estimate of the number of employees who left the employer during the current quarter. These individuals received UI-covered earnings during the current quarter but did not receive any UI-covered earnings in the next quarter from this employer. Separations are not defined for the last quarter of available data.

E.10 Individual concepts

Flow employment (m): for $qfirst \leq t \leq qlast$, individual i employed (matched to a job) at some time during period t at establishment j

$$m_{ijt} = \begin{cases} 1, & \text{if } i \text{ has positive earnings at establishment } j \text{ during quarter } t \\ 0, & \text{otherwise.} \end{cases} \tag{E.1}$$

Beginning of quarter employment (b): For $qfirst < t$, individual i employed at the end of $t-1$, beginning of t

$$b_{ijt} = \begin{cases} 1, & \text{if } m_{ijt-1} = m_{ijt} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{E.2}$$

End of quarter employment (e): For $t < qlast$, individual i employed at j at the end of t, beginning of $t+1$

$$e_{ijt} = \begin{cases} 1, & \text{if } m_{ijt} = m_{ijt+1} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{E.3}$$

Accessions (a_1): For $qfirst < t$, individual i acceded to j during t

$$a_{1ijt} = \begin{cases} 1, & \text{if } m_{ijt-1} = 0 \; \& \; m_{ijt} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{E.4}$$

Separations (s_1): For $t < qlast$, individual i separated from j during t

$$s_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt} = 1 \ \& \ m_{ijt+1} = 0 \\ 0, \text{ otherwise.} \end{cases} \tag{E.5}$$

Full quarter employment (f): For $qfirst < t < qlast$, individual i was employed at j at the beginning and end of quarter t (full-quarter job)

$$f_{ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = 1 \ \& \ m_{ijt} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{E.6}$$

New hires (h_1): For $qfirst+3 < t$, individual i was newly hired at j during period t

$$h_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt-4} = 0 \ \& \ m_{ijt-3} = 0 \ \& \ m_{ijt-2} = 0 \ \& \ m_{ijt-1} = 0 \ \& \ m_{ijt} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{E.7}$$

Recalls (r_1): For $qfirst + 3 < t$, individual$i$ was recalled from layoff at j during period t

$$r_{1ijt} = \begin{cases} 1, \text{ if } m_{ijt-1} = 0 \ \& \ m_{ijt} = 1 \ \& \ h_{ijt} = 0 \\ 0, \text{ otherwise.} \end{cases} \tag{E.8}$$

New hires to full quarter status (a_3): For $qfirst + 4 < t < qlast$, individual i transited from consecutive-quarter hired to full-quarter hired status at j at the start of $t+1$ (hired in $t-1$ and full-quarter employed in t)

$$h_{3ijt} = \begin{cases} 1, \text{ if } h_{2ijt-1} = 1 \ \& \ m_{ijt+1} = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{E.9}$$

Separations from full-quarter status (s_3): For $qfirst + 1 < t < qlast$, individual i separated from j during t with full-quarter status during $t-1$

$$s_{3ijt} = \begin{cases} 1, \text{ if } s_{2ijt} = 1 \ \& \ m_{ijt-2} = 1 \\ 0, \text{ otherwise} \end{cases} \tag{E.10}$$

Total earnings during the quarter (w_1): for $qfirst \leq t \leq qlast$, earnings of individual i at establishment j during period t

$$w_{1ijt} = \sum \text{all } UI \text{ covered earnings by } i \text{ at } j \text{ during } t \tag{E.11}$$

Earnings of end-of-period employees at establishment j during period t

$$w_{2ijt} = \begin{cases} w_{1ijt}, \text{ if } e_{ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{E.12}$$

Earnings of full-quarter individual i at establishment j during period t

$$w_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } f_{ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.13}$$

For $qfirst \leq t \leq qlast$, total earnings of individual i during period t

$$w_{1i\bullet t} = \sum_{j \text{ employs } i \text{ during } t} w_{1ijt} \tag{E.14}$$

Total earnings of end-of-period employees i during period t

$$w_{2i\bullet t} = \begin{cases} w_{1i\bullet t}, \text{ if } e_{ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.15}$$

Total earnings of full-quarter employees i during period t

$$w_{3i\bullet t} = \begin{cases} w_{1i\bullet t}, \text{ if } f_{ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.16}$$

For $qfirst < t$, change in total earnings of individual i between periods $t-1$ and t. The goal is to produce statistics based on:

$$\Delta w_{1i\bullet t} = w_{1i\bullet t} - w_{1i\bullet t-1} \tag{E.17}$$

Earnings of accessions to employer j during period t

$$wa_{1ijt} = \begin{cases} w_{1ijt}, \text{ if } a_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.18}$$

Earnings of full-quarter accessions to establishment j during period t

$$wa_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } a_{3ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.19}$$

Earnings of full-quarter new hires to establishment j during period t

$$wh_{3ijt} = \begin{cases} w_{1ijt}, \text{ if } h_{3ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.20}$$

Total earnings change for accessions to establishment j during t

$$\Delta wa_{1ijt} = \begin{cases} \Delta w_{1i\bullet t}, \text{ if } a_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.21}$$

Earnings of separations from establishment j during period t

$$ws_{1ijt} = \begin{cases} w_{1ijt}, \text{ if } s_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.22}$$

Total earnings change for separations from establishment j during t

$$\Delta ws_{1ijt} = \begin{cases} \Delta w_{1i\bullet t+1}, \text{ if } s_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.23}$$

Periods of non-employment prior to an accession by i at establishment j during t during the previous four quarters (defined for $qfirst + 3 < t$)

$$na_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } a_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.24}$$

where $n_{it} = 1$ if $m_{ijt} = 0 \ \forall j$.

Periods of non-employment prior to a new hire by i at establishment j during t during the previous four quarters

$$nh_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } h_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.25}$$

Periods of non-employment prior to a recall by i at establishment j during t during the previous four quarters

$$nr_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it-s}, \text{ if } r_{1ijt} = 1 \\ \\ \text{undefined, otherwise} \end{cases} \tag{E.26}$$

Periods of non-employment following a separation by i from establishment j during t during the next four quarters, (defined for $t < qlast - 3$)

$$ns_{ijt} = \begin{cases} \sum_{1 \leqslant s \leqslant 4} n_{it+s}, & \text{if } s_{1ijt} = 1 \\ \text{undefined, otherwise} \end{cases} \tag{E.27}$$

E.11 Establishment concepts

For statistic x_{cijt} denote the sum over i during period t as $x_{c \cdot jt}$. For example, beginning of period employment for firm j is written as:

$$b_{\cdot jt} = \sum_{i} b_{ijt} \tag{E.28}$$

All individual statistics generate establishment totals according to the formula above. The key establishment statistic is the average end-of-period employment growth rate for establishment j, the components of which are defined here.

Beginning-of-period employment (number of jobs)

$$B_{jt} = b_{\cdot jt} \tag{E.29}$$

End-of-period employment (number of jobs)

$$E_{jt} = e_{\cdot jt} \tag{E.30}$$

Employment any time during the period (number of jobs)

$$M_{jt} = m_{\cdot jt} \tag{E.31}$$

Full-quarter employment

$$F_{jt} = f_{\cdot jt} \tag{E.32}$$

Net job flows (change in employment) for establishment j during period t

$$JF_{jt} = E_{jt} - B_{jt} \tag{E.33}$$

Average employment for establishment j between periods $t-1$ and t

$$\bar{E}_{jt} = \frac{(B_{jt} + E_{jt})}{2} \tag{E.34}$$

Average employment growth rate for establishment j between periods $t-1$ and t

$$G_{jt} = \frac{JF_{jt}}{\bar{E}_{jt}} \tag{E.35}$$

Job creation for establishment j between periods $t-1$ and t

$$JC_{jt} = \bar{E}_{jt} \max{(0, G_{jt})} \tag{E.36}$$

Job destruction for establishment j between periods $t-1$ and t

$$JD_{jt} = \bar{E}_{jt} \,\mathrm{abs}\left(\min\left(0, G_{jt}\right)\right) \tag{E.37}$$

Net change in full-quarter employment for establishment j during period t

$$FJF_{jt} = F_{jt} - F_{jt-1} \tag{E.38}$$

Full-quarter job creations for establishment j between $t-1$ and t

$$FJC_{jt} = \bar{F}_{jt} \max\left(0, FG_{jt}\right) \tag{E.39}$$

Full-quarter job destruction for establishment j between $t-1$ and t

$$FJD_{jt} = \bar{F}_{jt} abs\left(\min\left(0, FG_{jt}\right)\right) \tag{E.40}$$

Accessions for establishment j during t

$$A_{jt} = a_{1\cdot jt} \tag{E.41}$$

Separations for establishment j during t

$$S_{jt} = s_{1\cdot jt} \tag{E.42}$$

New hires for establishment j during t

$$H_{jt} = h_{1 \cdot jt} \tag{E.43}$$

Full Quarter New hires   for establishment j during t

$$H_{3jt} = h_{3 \cdot jt} \tag{E.44}$$

Recalls   for establishment j during t

$$R_{jt} = r_{1 \cdot jt} \tag{E.45}$$

Flow into full-quarter employment   for establishment j during t

$$FA_{jt} = a_{3 \cdot jt} \tag{E.46}$$

New hires into full-quarter employment   for establishment j during t

$$FH_{jt} = h_{3 \cdot jt} \tag{E.47}$$

Flow out of full-quarter employment   for establishment j during t

$$FS_{jt} = s_{3 \cdot jt} \tag{E.48}$$

Total payroll of all employees

$$W_{1jt} = w_{1 \cdot jt} \tag{E.49}$$

Total payroll of end-of-period employees

$$W_{2jt} = w_{2\cdot jt} \tag{E.50}$$

Total payroll of full-quarter employees

$$W_{3jt} = w_{3\cdot jt} \tag{E.51}$$

Total payroll of accessions

$$WA_{jt} = wa_{1\cdot jt} \tag{E.52}$$

Change in total earnings for accessions

$$\Delta WA_{jt} = \sum_{i \in \{J(i,t)=j\}} \triangle wa_{1ijt} \tag{E.53}$$

Total payroll of transits to full-quarter status

$$WFA_{jt} = wa_{3\cdot jt} \tag{E.54}$$

Total payroll of new hires to full-quarter status

$$WFH_{jt} = wh_{3\cdot jt} \tag{E.55}$$

Total periods of non-employment for accessions

$$NA_{jt} = na_{\cdot jt} \tag{E.56}$$

Total periods of non-employment for new hires (last four quarters)

$$NH_{jt} = nh_{\cdot jt} \tag{E.57}$$

Total periods of non-employment for recalls (last four quarters)

$$NR_{jt} = nr_{\cdot jt} \tag{E.58}$$

Total earnings of separations

$$WS_{jt} = ws_{1 \cdot jt} \tag{E.59}$$

Total change in total earnings for separations

$$\Delta WS_{jt} = \sum_{i \in \{J(i,t)=j\}} \Delta ws_{1ijt} \tag{E.60}$$

Total earnings of separations from full-quarter status (most recent full quarter)

$$WFS_{jt} = ws_{3 \cdot jt} \tag{E.61}$$

Total change in total earnings for full-quarter separations

$$\Delta WFS_{jt} = \sum_{i \in \{J(i,t)=j\}} \Delta ws_{3ijt} \tag{E.62}$$

Total periods of non-employment for separations

$$NS_{jt} = ns_{\cdot jt} \tag{E.63}$$

Average earnings of end-of-period employees

$$ZW_{2jt} = W_{2jt} \,/\, E_{jt} \tag{E.64}$$

Average earnings of full-quarter employees

$$ZW_{3jt} = W_{3jt} \,/\, F_{jt} \tag{E.65}$$

Average earnings of accessions

$$ZWA_{jt} = WA_{jt} \,/\, A_{jt} \tag{E.66}$$

Average change in total earnings for accessions

$$Z\Delta WA_{jt} = \Delta WA_{jt} \,/\, A_{jt} \tag{E.67}$$

Average earnings of transits to full-quarter status

$$ZWFA_{jt} = WFA_{jt} \,/\, FA_{jt} \tag{E.68}$$

Average earnings of new hires to full-quarter status

$$ZWFH_{jt} = WFH_{jt} \,/\, FH_{jt} \tag{E.69}$$

Average periods of non-employment for accessions

$$ZNA_{jt} = NA_{jt} \,/\, A_{jt} \tag{E.70}$$

Average periods of non-employment for new hires (last four quarters)

$$ZNH_{jt} = NH_{jt} \,/\, H_{jt} \tag{E.71}$$

Average periods of non-employment for recalls (last four quarters)

$$ZNR_{jt} = NR_{jt} \,/\, R_{jt} \tag{E.72}$$

Average earnings of separations

$$ZWS_{jt} = WS_{jt} \,/\, S_{jt} \tag{E.73}$$

Average change in total earnings for separations

$$Z\Delta WS_{jt} = \Delta WS_{jt} \,/\, S_{jt} \tag{E.74}$$

Average earnings of separations from full-quarter status (most recent full quarter)

$$ZWFS_{jt-1} = WFS_{jt-1} \,/\, FS_{jt} \tag{E.75}$$

Average periods of non-employment for separations

$$ZNS_{jt} = NS_{jt} \,/\, S_{jt} \tag{E.76}$$

Appendix F

Bias Summary Statistics

Table F.1: Distribution of Bias Measures for County by SIC Division: M and B

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
M	Number of Obs	456100		456064		455994	
M	Mean	0.0700	-0.0002	0.0342	-0.0072	0.0327	-0.0071
M	Variance	14.9578	0.1388	0.1578	0.0221	0.1670	0.0221
M	Number of Missing Obs	42524		42576		42646	
M	99th Percentile	1.3333	0.3793	1.0000	0.3622	1.0000	0.3696
M	95th Percentile	0.4000	0.1833	0.3929	0.1736	0.3846	0.1764
M	90th Percentile	0.2222	0.1145	0.2204	0.1059	0.2143	0.1081
M	75th Percentile	0.0714	0.0433	0.0652	0.0353	0.0631	0.0350
M	50th Percentile	0.0000	0.0000	0.0000	-0.0053	0.0000	-0.0056
M	25th Percentile	-0.0357	-0.0289	-0.0417	-0.0352	-0.0428	-0.0357
M	10th Percentile	-0.1818	-0.0966	-0.1667	-0.1001	-0.1739	-0.1034
M	5th Percentile	-0.3333	-0.1970	-0.3077	-0.1938	-0.3158	-0.1990
M	1st Percentile	-0.9313	-0.6676	-0.6829	-0.6488	-0.6914	-0.6321
M	90th - 10th	0.4040	0.2111	0.3871	0.2061	0.3882	0.2116
M	75th - 25th	0.1071	0.0722	0.1069	0.0705	0.1059	0.0707
B	Number of Obs	435648		435565		435504	
B	Mean	0.0667	0.0071	0.0370	0.0000	0.0350	0.0001
B	Variance	8.0617	0.1217	0.2003	0.0231	0.1987	0.0232
B	Number of Missing Obs	62976		63075		63136	
B	99th Percentile	1.2917	0.3805	1.0000	0.3670	1.0000	0.3708
B	95th Percentile	0.4000	0.1861	0.4000	0.1756	0.4000	0.1800
B	90th Percentile	0.2245	0.1183	0.2258	0.1097	0.2222	0.1111
B	75th Percentile	0.0702	0.0486	0.0667	0.0406	0.0645	0.0402
B	50th Percentile	0.0000	0.0126	0.0000	0.0043	0.0000	0.0043
B	25th Percentile	-0.0324	-0.0183	-0.0385	-0.0247	-0.0400	-0.0256
B	10th Percentile	-0.1852	-0.0898	-0.1714	-0.0934	-0.1770	-0.0965
B	5th Percentile	-0.3360	-0.1902	-0.3192	-0.1891	-0.3333	-0.1933
B	1st Percentile	-0.9965	-0.6851	-0.7044	-0.6662	-0.7135	-0.6485
B	90th - 10th	0.4097	0.2081	0.3972	0.2031	0.3993	0.2076
B	75th - 25th	0.1026	0.0669	0.1051	0.0653	0.1045	0.0658

Table F.2: Distribution of Bias Measures for County by SIC Division: E and F

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
E	Number of Obs	436115		436060		436003	
E	Mean	0.0652	0.0062	0.0351	-0.0010	0.0330	-0.0009
E	Variance	14.4526	0.1541	0.1790	0.0228	0.1846	0.0229
E	Number of Missing Obs	62509		62580		62637	
E	99th Percentile	1.2222	0.3795	1.0000	0.3642	1.0000	0.3699
E	95th Percentile	0.4000	0.1834	0.4000	0.1738	0.4000	0.1765
E	90th Percentile	0.2222	0.1172	0.2222	0.1085	0.2209	0.1103
E	75th Percentile	0.0681	0.0471	0.0650	0.0394	0.0625	0.0391
E	50th Percentile	0.0000	0.0113	0.0000	0.0033	0.0000	0.0030
E	25th Percentile	-0.0333	-0.0180	-0.0399	-0.0250	-0.0417	-0.0256
E	10th Percentile	-0.1858	-0.0894	-0.1739	-0.0934	-0.1787	-0.0963
E	5th Percentile	-0.3333	-0.1923	-0.3221	-0.1912	-0.3333	-0.1959
E	1st Percentile	-0.9859	-0.6815	-0.7066	-0.6636	-0.7129	-0.6456
E	90th - 10th	0.4080	0.2066	0.3961	0.2019	0.3996	0.2066
E	75th - 25th	0.1014	0.0650	0.1048	0.0644	0.1042	0.0647
F	Number of Obs	416691		416702		416624	
F	Mean	0.0637	0.0100	0.0371	0.0026	0.0347	0.0028
F	Variance	8.2222	0.1337	0.2352	0.0242	0.2381	0.0244
F	Number of Missing Obs	81933		81938		82016	
F	99th Percentile	1.1875	0.3896	1.0000	0.3750	1.0000	0.3755
F	95th Percentile	0.4000	0.1864	0.4186	0.1764	0.4087	0.1806
F	90th Percentile	0.2222	0.1207	0.2308	0.1111	0.2231	0.1132
F	75th Percentile	0.0672	0.0513	0.0656	0.0440	0.0627	0.0434
F	50th Percentile	0.0000	0.0158	0.0000	0.0078	0.0000	0.0077
F	25th Percentile	-0.0308	-0.0139	-0.0376	-0.0210	-0.0397	-0.0217
F	10th Percentile	-0.1901	-0.0857	-0.1765	-0.0899	-0.1818	-0.0932
F	5th Percentile	-0.3448	-0.1900	-0.3333	-0.1879	-0.3333	-0.1915
F	1st Percentile	-1.0000	-0.6962	-0.7213	-0.6746	-0.7304	-0.6618
F	90th - 10th	0.4123	0.2064	0.4072	0.2010	0.4049	0.2064
F	75th - 25th	0.0980	0.0652	0.1032	0.0649	0.1024	0.0651

Table F.3: Distribution of Bias Measures for County by SIC Division: W1 and Z_NA

Relevant Variable	Statistic	Raw No Establishment Weight Applied Unweighted Bias	Raw No Establishment Weight Applied Weighted Bias	Raw Establishment Weight Applied Unweighted Bias	Raw Establishment Weight Applied Weighted Bias	Fuzzed Establishment Weight Applied Unweighted Bias	Fuzzed Establishment Weight Applied Weighted Bias
W1	Number of Obs	457153		456778		456778	
W1	Mean	0.3576	0.0072	0.1545	-0.0008	0.1593	-0.0005
W1	Variance	1592.2613	7.3273	121.0155	0.5696	138.4870	0.6476
W1	Number of Missing Obs	41471		41862		41862	
W1	99th Percentile	2.9573	0.4687	2.2481	0.4514	2.2681	0.4607
W1	95th Percentile	0.5880	0.2197	0.5340	0.2108	0.5390	0.2157
W1	90th Percentile	0.3058	0.1376	0.2845	0.1283	0.2861	0.1316
W1	75th Percentile	0.0965	0.0446	0.0856	0.0377	0.0856	0.0370
W1	50th Percentile	0.0000	0.0013	-0.0037	-0.0056	-0.0036	-0.0062
W1	25th Percentile	-0.0388	-0.0269	-0.0417	-0.0337	-0.0424	-0.0348
W1	10th Percentile	-0.2088	-0.0976	-0.1987	-0.1009	-0.2001	-0.1031
W1	5th Percentile	-0.3994	-0.2054	-0.3599	-0.2044	-0.3623	-0.2063
W1	1st Percentile	-0.8759	-0.7230	-0.7442	-0.7066	-0.7466	-0.6884
W1	90th - 10th	0.5146	0.2351	0.4831	0.2292	0.4862	0.2348
W1	75th - 25th	0.1353	0.0715	0.1273	0.0714	0.1280	0.0718
Z_NA	Number of Obs	317988		317759		317759	
Z_NA	Mean	0.0298	0.0223	0.0322	0.0221	0.0338	0.0223
Z_NA	Variance	0.1171	0.0301	0.1093	0.0296	0.1155	0.0302
Z_NA	Number of Missing Obs	180636		180881		180881	
Z_NA	99th Percentile	1.1642	0.4457	1.1379	0.4393	1.1741	0.4499
Z_NA	95th Percentile	0.3900	0.1579	0.3793	0.1571	0.3873	0.1653
Z_NA	90th Percentile	0.2083	0.0894	0.2011	0.0887	0.2073	0.0925
Z_NA	75th Percentile	0.0517	0.0348	0.0500	0.0349	0.0516	0.0351
Z_NA	50th Percentile	0.0000	0.0097	0.0000	0.0094	0.0000	0.0096
Z_NA	25th Percentile	-0.0350	-0.0127	-0.0326	-0.0131	-0.0337	-0.0131
Z_NA	10th Percentile	-0.1466	-0.0476	-0.1374	-0.0474	-0.1395	-0.0496
Z_NA	5th Percentile	-0.2727	-0.0882	-0.2500	-0.0866	-0.2531	-0.0912
Z_NA	1st Percentile	-1.0000	-0.2339	-0.7215	-0.2252	-0.7212	-0.2302
Z_NA	90th - 10th	0.3549	0.1370	0.3385	0.1361	0.3468	0.1421
Z_NA	75th - 25th	0.0867	0.0475	0.0826	0.0480	0.0853	0.0482

Table F.4: Distribution of Bias Measures for County by SIC Division: Z_NH and Z_NR

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NH	Number of Obs	292418		292198		292198	
Z_NH	Mean	0.0316	0.0279	0.0337	0.0277	0.0355	0.0280
Z_NH	Variance	0.1410	0.0424	0.1338	0.0418	0.1399	0.0420
Z_NH	Number of Missing Obs	206206		206442		206442	
Z_NH	99th Percentile	1.2692	0.4697	1.2599	0.4830	1.2857	0.4891
Z_NH	95th Percentile	0.4211	0.1678	0.4091	0.1667	0.4194	0.1753
Z_NH	90th Percentile	0.2207	0.0970	0.2143	0.0965	0.2208	0.1007
Z_NH	75th Percentile	0.0532	0.0405	0.0511	0.0410	0.0535	0.0412
Z_NH	50th Percentile	0.0000	0.0149	0.0000	0.0147	0.0000	0.0135
Z_NH	25th Percentile	-0.0400	-0.0090	-0.0374	-0.0095	-0.0385	-0.0106
Z_NH	10th Percentile	-0.1608	-0.0474	-0.1511	-0.0472	-0.1530	-0.0502
Z_NH	5th Percentile	-0.2982	-0.0897	-0.2775	-0.0878	-0.2805	-0.0931
Z_NH	1st Percentile	-1.0000	-0.2523	-0.8657	-0.2491	-0.8652	-0.2455
Z_NH	90th - 10th	0.3815	0.1444	0.3654	0.1437	0.3738	0.1509
Z_NH	75th - 25th	0.0932	0.0495	0.0885	0.0505	0.0920	0.0518
Z_NR	Number of Obs	207648		207570		207570	
Z_NR	Mean	0.0012	0.0144	0.0040	0.0135	0.0051	0.0140
Z_NR	Variance	0.0773	0.0436	0.0709	0.0366	0.0744	0.0385
Z_NR	Number of Missing Obs	290976		291070		291070	
Z_NR	99th Percentile	1.0000	0.7031	0.9200	0.6200	0.9345	0.6404
Z_NR	95th Percentile	0.3158	0.1940	0.3023	0.1839	0.3077	0.1904
Z_NR	90th Percentile	0.1437	0.0952	0.1400	0.0941	0.1429	0.0953
Z_NR	75th Percentile	0.0060	0.0256	0.0065	0.0253	0.0100	0.0261
Z_NR	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_NR	25th Percentile	-0.0241	-0.0236	-0.0217	-0.0231	-0.0231	-0.0225
Z_NR	10th Percentile	-0.1469	-0.0737	-0.1382	-0.0714	-0.1395	-0.0746
Z_NR	5th Percentile	-0.2967	-0.1329	-0.2733	-0.1277	-0.2768	-0.1319
Z_NR	1st Percentile	-1.0000	-0.3773	-1.0000	-0.3583	-1.0000	-0.3737
Z_NR	90th - 10th	0.2906	0.1689	0.2782	0.1655	0.2824	0.1699
Z_NR	75th - 25th	0.0301	0.0493	0.0282	0.0484	0.0331	0.0486

Table F.5: Distribution of Bias Measures for County by SIC Division: Z_NS and Z_W2

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NS	Number of Obs	318195		318000		318000	
Z_NS	Mean	0.0246	0.0249	0.0277	0.0242	0.0293	0.0245
Z_NS	Variance	0.1066	0.0231	0.0991	0.0217	0.1047	0.0234
Z_NS	Number of Missing Obs	180429		180640		180640	
Z_NS	99th Percentile	1.1000	0.4167	1.0906	0.3984	1.1113	0.4156
Z_NS	95th Percentile	0.3720	0.1489	0.3650	0.1460	0.3748	0.1520
Z_NS	90th Percentile	0.1976	0.0902	0.1917	0.0885	0.1978	0.0913
Z_NS	75th Percentile	0.0500	0.0414	0.0485	0.0403	0.0502	0.0400
Z_NS	50th Percentile	0.0000	0.0172	0.0000	0.0168	0.0000	0.0162
Z_NS	25th Percentile	-0.0365	-0.0078	-0.0339	-0.0080	-0.0348	-0.0085
Z_NS	10th Percentile	-0.1496	-0.0441	-0.1398	-0.0438	-0.1419	-0.0460
Z_NS	5th Percentile	-0.2772	-0.0855	-0.2535	-0.0833	-0.2570	-0.0870
Z_NS	1st Percentile	-1.0000	-0.2305	-0.7500	-0.2183	-0.7533	-0.2282
Z_NS	90th - 10th	0.3472	0.1343	0.3315	0.1323	0.3397	0.1373
Z_NS	75th - 25th	0.0865	0.0492	0.0824	0.0483	0.0850	0.0485
Z_W2	Number of Obs	436934		436633		436633	
Z_W2	Mean	0.0512	-0.0016	0.0473	-0.0009	0.0485	-0.0008
Z_W2	Variance	1.8981	0.0184	1.0326	0.0123	1.0342	0.0130
Z_W2	Number of Missing Obs	61690		62007		62007	
Z_W2	99th Percentile	1.0581	0.2047	0.9486	0.2024	0.9760	0.2158
Z_W2	95th Percentile	0.2596	0.0811	0.2455	0.0797	0.2530	0.0864
Z_W2	90th Percentile	0.1316	0.0467	0.1266	0.0461	0.1296	0.0495
Z_W2	75th Percentile	0.0355	0.0125	0.0344	0.0122	0.0349	0.0125
Z_W2	50th Percentile	0.0000	-0.0050	0.0000	-0.0046	0.0000	-0.0047
Z_W2	25th Percentile	-0.0250	-0.0181	-0.0232	-0.0174	-0.0241	-0.0184
Z_W2	10th Percentile	-0.1007	-0.0432	-0.0931	-0.0421	-0.0955	-0.0445
Z_W2	5th Percentile	-0.1932	-0.0743	-0.1728	-0.0726	-0.1768	-0.0759
Z_W2	1st Percentile	-0.5820	-0.2444	-0.4199	-0.2342	-0.4274	-0.2355
Z_W2	90th - 10th	0.2323	0.0899	0.2197	0.0882	0.2250	0.0940
Z_W2	75th - 25th	0.0604	0.0306	0.0576	0.0295	0.0590	0.0309

Table F.6: Distribution of Bias Measures for County by SIC Division: Z_W3 and Z_WFA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_W3	Number of Obs	417624		417371		417371	
Z_W3	Mean	0.0459	-0.0043	0.0434	-0.0037	0.0447	-0.0034
Z_W3	Variance	1.8939	0.0196	1.0138	0.0128	1.0225	0.0135
Z_W3	Number of Missing Obs	81000		81269		81269	
Z_W3	99th Percentile	0.9666	0.1934	0.8927	0.1887	0.9179	0.2003
Z_W3	95th Percentile	0.2466	0.0770	0.2355	0.0761	0.2419	0.0826
Z_W3	90th Percentile	0.1246	0.0434	0.1208	0.0430	0.1238	0.0465
Z_W3	75th Percentile	0.0330	0.0100	0.0321	0.0098	0.0327	0.0106
Z_W3	50th Percentile	0.0000	-0.0069	0.0000	-0.0067	0.0000	-0.0067
Z_W3	25th Percentile	-0.0236	-0.0214	-0.0221	-0.0209	-0.0232	-0.0216
Z_W3	10th Percentile	-0.0959	-0.0461	-0.0892	-0.0451	-0.0917	-0.0477
Z_W3	5th Percentile	-0.1858	-0.0744	-0.1669	-0.0725	-0.1710	-0.0757
Z_W3	1st Percentile	-0.5762	-0.2462	-0.4152	-0.2342	-0.4234	-0.2388
Z_W3	90th - 10th	0.2204	0.0895	0.2100	0.0881	0.2155	0.0942
Z_W3	75th - 25th	0.0566	0.0314	0.0542	0.0307	0.0559	0.0322
Z_WFA	Number of Obs	291739		291613		291613	
Z_WFA	Mean	0.1710	-0.0151	0.1755	-0.0145	0.1794	-0.0136
Z_WFA	Variance	18.4289	0.7112	18.7959	0.6812	16.9221	0.6266
Z_WFA	Number of Missing Obs	206885		207027		207027	
Z_WFA	99th Percentile	2.9100	0.4780	2.9257	0.4721	3.0291	0.4791
Z_WFA	95th Percentile	0.5694	0.1446	0.5640	0.1455	0.5797	0.1486
Z_WFA	90th Percentile	0.2540	0.0693	0.2516	0.0698	0.2581	0.0734
Z_WFA	75th Percentile	0.0474	0.0061	0.0467	0.0065	0.0482	0.0072
Z_WFA	50th Percentile	0.0000	-0.0276	0.0000	-0.0263	0.0000	-0.0273
Z_WFA	25th Percentile	-0.0424	-0.0669	-0.0405	-0.0651	-0.0412	-0.0663
Z_WFA	10th Percentile	-0.1654	-0.1143	-0.1580	-0.1129	-0.1596	-0.1141
Z_WFA	5th Percentile	-0.3054	-0.1675	-0.2872	-0.1660	-0.2902	-0.1636
Z_WFA	1st Percentile	-1.0000	-0.4222	-0.8393	-0.4207	-0.8413	-0.3983
Z_WFA	90th - 10th	0.4194	0.1836	0.4097	0.1827	0.4177	0.1875
Z_WFA	75th - 25th	0.0897	0.0730	0.0871	0.0716	0.0894	0.0735

Table F.7: Distribution of Bias Measures for County by SIC Division: Z_WFS and Z_WH3

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_WFS	Number of Obs	279139		279033		279033	
Z_WFS	Mean	0.4611	0.0403	0.4643	0.0400	0.4903	0.0418
Z_WFS	Variance	1575.2812	66.3155	1481.3692	59.3746	2210.0538	88.2057
Z_WFS	Number of Missing Obs	219485		219607		219607	
Z_WFS	99th Percentile	6.0171	0.7801	6.0435	0.7649	6.2045	0.8151
Z_WFS	95th Percentile	0.8941	0.2436	0.8868	0.2413	0.9000	0.2462
Z_WFS	90th Percentile	0.3758	0.1347	0.3714	0.1328	0.3778	0.1341
Z_WFS	75th Percentile	0.0710	0.0401	0.0702	0.0398	0.0707	0.0404
Z_WFS	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_WFS	25th Percentile	-0.0363	-0.0336	-0.0345	-0.0338	-0.0349	-0.0351
Z_WFS	10th Percentile	-0.1881	-0.1017	-0.1801	-0.1000	-0.1818	-0.1015
Z_WFS	5th Percentile	-0.3766	-0.1746	-0.3541	-0.1703	-0.3563	-0.1732
Z_WFS	1st Percentile	-1.0000	-0.4527	-1.0000	-0.4305	-1.0000	-0.4357
Z_WFS	90th - 10th	0.5639	0.2363	0.5515	0.2328	0.5597	0.2355
Z_WFS	75th - 25th	0.1072	0.0737	0.1047	0.0736	0.1056	0.0755
Z_WH3	Number of Obs	248824		248725		248725	
Z_WH3	Mean	0.1599	-0.0195	0.1668	-0.0187	0.1716	-0.0177
Z_WH3	Variance	15.5682	0.6838	17.6954	0.7176	16.4230	0.6938
Z_WH3	Number of Missing Obs	249800		249915		249915	
Z_WH3	99th Percentile	2.9271	0.4643	2.9556	0.4647	3.0435	0.4649
Z_WH3	95th Percentile	0.5842	0.1479	0.5812	0.1503	0.5961	0.1557
Z_WH3	90th Percentile	0.2651	0.0717	0.2641	0.0720	0.2717	0.0760
Z_WH3	75th Percentile	0.0508	0.0051	0.0504	0.0060	0.0520	0.0068
Z_WH3	50th Percentile	0.0000	-0.0308	0.0000	-0.0299	0.0000	-0.0309
Z_WH3	25th Percentile	-0.0478	-0.0735	-0.0463	-0.0719	-0.0470	-0.0729
Z_WH3	10th Percentile	-0.1768	-0.1239	-0.1701	-0.1221	-0.1727	-0.1221
Z_WH3	5th Percentile	-0.3199	-0.1832	-0.3036	-0.1818	-0.3072	-0.1785
Z_WH3	1st Percentile	-1.0000	-0.4417	-0.9725	-0.4478	-0.9737	-0.4279
Z_WH3	90th - 10th	0.4420	0.1956	0.4342	0.1942	0.4445	0.1981
Z_WH3	75th - 25th	0.0986	0.0786	0.0967	0.0779	0.0990	0.0797

Table F.8: Distribution of Bias Measures for County by SIC Division: A and FA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
A	Number of Obs	486496	n/a	486512	n/a	486512	n/a
A	Mean	-1.2712	n/a	-1.5837	n/a	-1.6025	n/a
A	Variance	985.3309	n/a	1238.4066	n/a	1279.2886	n/a
A	Number of Missing Obs	12128	n/a	12128	n/a	12128	n/a
A	99th Percentile	34.0000	n/a	34.0000	n/a	34.0000	n/a
A	95th Percentile	6.0000	n/a	6.0000	n/a	6.0000	n/a
A	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
A	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
A	5th Percentile	-6.0000	n/a	-7.0000	n/a	-7.0000	n/a
A	1st Percentile	-54.0000	n/a	-59.0000	n/a	-60.0000	n/a
A	90th - 10th	4.0000	n/a	4.0000	n/a	4.0000	n/a
A	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	Number of Obs	462208	n/a	462224	n/a	462224	n/a
FA	Mean	1.0809	n/a	1.0112	n/a	1.0217	n/a
FA	Variance	374.8755	n/a	415.6078	n/a	427.0300	n/a
FA	Number of Missing Obs	36416	n/a	36416	n/a	36416	n/a
FA	99th Percentile	38.0000	n/a	38.0000	n/a	38.0000	n/a
FA	95th Percentile	6.0000	n/a	6.0000	n/a	6.0000	n/a
FA	90th Percentile	2.0000	n/a	3.0000	n/a	3.0000	n/a
FA	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FA	5th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
FA	1st Percentile	-15.0000	n/a	-17.0000	n/a	-17.0000	n/a
FA	90th - 10th	3.0000	n/a	4.0000	n/a	4.0000	n/a
FA	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.9: Distribution of Bias Measures for County by SIC Division: FJC and FJD

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJC	Number of Obs	462208	n/a	462224	n/a	420306	n/a
FJC	Mean	0.6883	n/a	0.6269	n/a	0.6814	n/a
FJC	Variance	339.9791	n/a	388.3463	n/a	425.3687	n/a
FJC	Number of Missing Obs	36416	n/a	36416	n/a	78334	n/a
FJC	99th Percentile	28.0000	n/a	28.0000	n/a	31.0000	n/a
FJC	95th Percentile	5.0000	n/a	5.0000	n/a	5.0000	n/a
FJC	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
FJC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FJC	5th Percentile	-2.0000	n/a	-2.0000	n/a	-3.0000	n/a
FJC	1st Percentile	-13.0000	n/a	-14.0000	n/a	-16.0000	n/a
FJC	90th - 10th	3.0000	n/a	3.0000	n/a	3.0000	n/a
FJC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	Number of Obs	462208	n/a	462224	n/a	420306	n/a
FJD	Mean	-0.8377	n/a	-0.9195	n/a	-1.0184	n/a
FJD	Variance	143.0356	n/a	157.1704	n/a	173.4379	n/a
FJD	Number of Missing Obs	36416	n/a	36416	n/a	78334	n/a
FJD	99th Percentile	5.0000	n/a	5.0000	n/a	6.0000	n/a
FJD	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FJD	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FJD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	10th Percentile	-1.0000	n/a	-2.0000	n/a	-2.0000	n/a
FJD	5th Percentile	-3.0000	n/a	-4.0000	n/a	-4.0000	n/a
FJD	1st Percentile	-22.0000	n/a	-23.0000	n/a	-26.0000	n/a
FJD	90th - 10th	2.0000	n/a	3.0000	n/a	3.0000	n/a
FJD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.10: Distribution of Bias Measures for County by SIC Division: FJF and FS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJF	Number of Obs	462208	n/a	462224	n/a	420306	n/a
FJF	Mean	1.5227	n/a	1.5450	n/a	1.6979	n/a
FJF	Variance	580.0201	n/a	644.7980	n/a	706.4966	n/a
FJF	Number of Missing Obs	36416	n/a	36416	n/a	78334	n/a
FJF	99th Percentile	43.0000	n/a	45.0000	n/a	49.0000	n/a
FJF	95th Percentile	7.0000	n/a	8.0000	n/a	9.0000	n/a
FJF	90th Percentile	3.0000	n/a	3.0000	n/a	4.0000	n/a
FJF	75th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FJF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FJF	5th Percentile	-2.0000	n/a	-3.0000	n/a	-3.0000	n/a
FJF	1st Percentile	-14.0000	n/a	-15.0000	n/a	-17.0000	n/a
FJF	90th - 10th	4.0000	n/a	4.0000	n/a	5.0000	n/a
FJF	75th - 25th	1.0000	n/a	1.0000	n/a	1.0000	n/a
FS	Number of Obs	462208	n/a	462224	n/a	462224	n/a
FS	Mean	-0.4468	n/a	-0.5345	n/a	-0.5407	n/a
FS	Variance	126.4787	n/a	140.5689	n/a	144.0999	n/a
FS	Number of Missing Obs	36416	n/a	36416	n/a	36416	n/a
FS	99th Percentile	11.0000	n/a	11.0000	n/a	11.0000	n/a
FS	95th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
FS	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FS	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FS	5th Percentile	-3.0000	n/a	-3.0000	n/a	-4.0000	n/a
FS	1st Percentile	-21.0000	n/a	-23.0000	n/a	-23.0000	n/a
FS	90th - 10th	2.0000	n/a	2.0000	n/a	2.0000	n/a
FS	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.11: Distribution of Bias Measures for County by SIC Division: FT and H

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FT	Number of Obs	460928	n/a	460480	n/a	460480	n/a
FT	Mean	0.0097	n/a	0.0101	n/a	0.0102	n/a
FT	Variance	0.1006	n/a	0.0808	n/a	0.0949	n/a
FT	Number of Missing Obs	37696	n/a	38160	n/a	38160	n/a
FT	99th Percentile	0.4000	n/a	0.3100	n/a	0.3140	n/a
FT	95th Percentile	0.0700	n/a	0.0600	n/a	0.0630	n/a
FT	90th Percentile	0.0300	n/a	0.0300	n/a	0.0320	n/a
FT	75th Percentile	0.0100	n/a	0.0100	n/a	0.0090	n/a
FT	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	25th Percentile	0.0000	n/a	0.0000	n/a	-0.0030	n/a
FT	10th Percentile	-0.0300	n/a	-0.0300	n/a	-0.0260	n/a
FT	5th Percentile	-0.0600	n/a	-0.0600	n/a	-0.0570	n/a
FT	1st Percentile	-0.2700	n/a	-0.2200	n/a	-0.2240	n/a
FT	90th - 10th	0.0600	n/a	0.0600	n/a	0.0580	n/a
FT	75th - 25th	0.0100	n/a	0.0100	n/a	0.0120	n/a
H	Number of Obs	450080	n/a	450096	n/a	450096	n/a
H	Mean	-1.3022	n/a	-1.5732	n/a	-1.5929	n/a
H	Variance	888.4246	n/a	1119.4312	n/a	1158.5343	n/a
H	Number of Missing Obs	48544	n/a	48544	n/a	48544	n/a
H	99th Percentile	29.0000	n/a	28.0000	n/a	28.0000	n/a
H	95th Percentile	5.0000	n/a	5.0000	n/a	5.0000	n/a
H	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
H	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
H	5th Percentile	-5.0000	n/a	-6.0000	n/a	-6.0000	n/a
H	1st Percentile	-48.0000	n/a	-54.0000	n/a	-54.0000	n/a
H	90th - 10th	4.0000	n/a	4.0000	n/a	4.0000	n/a
H	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.12: Distribution of Bias Measures for County by SIC Division: H3 and JC

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
H3	Number of Obs	425728	n/a	425744	n/a	425744	n/a
H3	Mean	1.0562	n/a	1.0134	n/a	1.0220	n/a
H3	Variance	339.7536	n/a	383.5239	n/a	394.5597	n/a
H3	Number of Missing Obs	72896	n/a	72896	n/a	72896	n/a
H3	99th Percentile	36.0000	n/a	36.0000	n/a	36.0000	n/a
H3	95th Percentile	6.0000	n/a	6.0000	n/a	6.0000	n/a
H3	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
H3	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
H3	5th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
H3	1st Percentile	-13.0000	n/a	-15.0000	n/a	-15.0000	n/a
H3	90th - 10th	3.0000	n/a	3.0000	n/a	3.0000	n/a
H3	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	Number of Obs	474336	n/a	474352	n/a	439269	n/a
JC	Mean	-0.7118	n/a	-0.8482	n/a	-0.9223	n/a
JC	Variance	234.6285	n/a	329.0003	n/a	358.0168	n/a
JC	Number of Missing Obs	24288	n/a	24288	n/a	59371	n/a
JC	99th Percentile	12.0000	n/a	12.0000	n/a	13.0000	n/a
JC	95th Percentile	2.0000	n/a	3.0000	n/a	3.0000	n/a
JC	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
JC	5th Percentile	-4.0000	n/a	-4.0000	n/a	-5.0000	n/a
JC	1st Percentile	-24.0000	n/a	-27.0000	n/a	-29.0000	n/a
JC	90th - 10th	3.0000	n/a	3.0000	n/a	3.0000	n/a
JC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.13: Distribution of Bias Measures for County by SIC Division: JD and JF

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
JD	Number of Obs	474336	n/a	474352	n/a	439269	n/a
JD	Mean	-0.5679	n/a	-0.6687	n/a	-0.7295	n/a
JD	Variance	177.8145	n/a	190.9215	n/a	206.3855	n/a
JD	Number of Missing Obs	24288	n/a	24288	n/a	59371	n/a
JD	99th Percentile	11.0000	n/a	11.0000	n/a	11.0000	n/a
JD	95th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
JD	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
JD	5th Percentile	-4.0000	n/a	-4.0000	n/a	-5.0000	n/a
JD	1st Percentile	-22.0000	n/a	-23.0000	n/a	-25.0000	n/a
JD	90th - 10th	3.0000	n/a	3.0000	n/a	3.0000	n/a
JD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	Number of Obs	474336	n/a	474352	n/a	439269	n/a
JF	Mean	-0.1439	n/a	-0.1778	n/a	-0.1920	n/a
JF	Variance	289.2791	n/a	367.6756	n/a	398.0128	n/a
JF	Number of Missing Obs	24288	n/a	24288	n/a	59371	n/a
JF	99th Percentile	24.0000	n/a	25.0000	n/a	27.0000	n/a
JF	95th Percentile	5.0000	n/a	5.0000	n/a	6.0000	n/a
JF	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
JF	75th Percentile	0.0000	n/a	0.0000	n/a	1.0000	n/a
JF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
JF	5th Percentile	-5.0000	n/a	-5.0000	n/a	-6.0000	n/a
JF	1st Percentile	-26.0000	n/a	-28.0000	n/a	-30.0000	n/a
JF	90th - 10th	4.0000	n/a	4.0000	n/a	4.0000	n/a
JF	75th - 25th	0.0000	n/a	0.0000	n/a	1.0000	n/a

Table F.14: Distribution of Bias Measures for County by SIC Division: R and S

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
R	Number of Obs	450080	n/a	450096	n/a	450096	n/a
R	Mean	-0.0355	n/a	-0.0613	n/a	-0.0684	n/a
R	Variance	45.8843	n/a	53.1354	n/a	53.0522	n/a
R	Number of Missing Obs	48544	n/a	48544	n/a	48544	n/a
R	99th Percentile	9.0000	n/a	9.0000	n/a	9.0000	n/a
R	95th Percentile	1.0000	n/a	2.0000	n/a	2.0000	n/a
R	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
R	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	10th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
R	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
R	1st Percentile	-8.0000	n/a	-9.0000	n/a	-9.0000	n/a
R	90th - 10th	2.0000	n/a	2.0000	n/a	2.0000	n/a
R	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	Number of Obs	486464	n/a	486480	n/a	486480	n/a
S	Mean	-1.1256	n/a	-1.4016	n/a	-1.4112	n/a
S	Variance	778.2205	n/a	899.6368	n/a	926.3721	n/a
S	Number of Missing Obs	12160	n/a	12160	n/a	12160	n/a
S	99th Percentile	32.0000	n/a	30.0000	n/a	31.0000	n/a
S	95th Percentile	6.0000	n/a	6.0000	n/a	6.0000	n/a
S	90th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
S	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	10th Percentile	-2.0000	n/a	-2.0000	n/a	-2.0000	n/a
S	5th Percentile	-6.0000	n/a	-7.0000	n/a	-7.0000	n/a
S	1st Percentile	-51.0000	n/a	-55.0000	n/a	-56.0000	n/a
S	90th - 10th	4.0000	n/a	4.0000	n/a	4.0000	n/a
S	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.15: Distribution of Bias Measures for County by SIC Division: Z_dWA and Z_dWS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_dWA	Number of Obs	391709	n/a	389986	n/a	389986	n/a
Z_dWA	Mean	81.1152	n/a	76.4721	n/a	77.9429	n/a
Z_dWA	Variance	260855965.2690	n/a	210277710.2838	n/a	229445392.9054	n/a
Z_dWA	Number of Missing Obs	106915	n/a	108654	n/a	108654	n/a
Z_dWA	99th Percentile	2203.0000	n/a	2084.0000	n/a	2081.0000	n/a
Z_dWA	95th Percentile	604.0000	n/a	561.0000	n/a	561.0000	n/a
Z_dWA	90th Percentile	268.0000	n/a	252.0000	n/a	252.0000	n/a
Z_dWA	75th Percentile	51.0000	n/a	48.0000	n/a	48.0000	n/a
Z_dWA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWA	25th Percentile	-24.0000	n/a	-22.0000	n/a	-22.0000	n/a
Z_dWA	10th Percentile	-163.0000	n/a	-150.0000	n/a	-150.0000	n/a
Z_dWA	5th Percentile	-378.0000	n/a	-345.0000	n/a	-346.0000	n/a
Z_dWA	1st Percentile	-1537.0000	n/a	-1361.0000	n/a	-1374.0000	n/a
Z_dWA	90th - 10th	431.0000	n/a	402.0000	n/a	402.0000	n/a
Z_dWA	75th - 25th	75.0000	n/a	70.0000	n/a	70.0000	n/a
Z_dWS	Number of Obs	391430	n/a	389754	n/a	389754	n/a
Z_dWS	Mean	-80.6299	n/a	-80.4773	n/a	-83.7519	n/a
Z_dWS	Variance	153394532.2143	n/a	147321518.7881	n/a	181898615.0364	n/a
Z_dWS	Number of Missing Obs	107194	n/a	108886	n/a	108886	n/a
Z_dWS	99th Percentile	1647.0000	n/a	1463.0000	n/a	1469.0000	n/a
Z_dWS	95th Percentile	404.0000	n/a	371.0000	n/a	370.0000	n/a
Z_dWS	90th Percentile	174.0000	n/a	162.0000	n/a	162.0000	n/a
Z_dWS	75th Percentile	27.0000	n/a	25.0000	n/a	25.0000	n/a
Z_dWS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWS	25th Percentile	-53.0000	n/a	-50.0000	n/a	-51.0000	n/a
Z_dWS	10th Percentile	-273.0000	n/a	-257.0000	n/a	-257.0000	n/a
Z_dWS	5th Percentile	-620.0000	n/a	-580.0000	n/a	-583.0000	n/a
Z_dWS	1st Percentile	-2306.0000	n/a	-2196.0000	n/a	-2233.0000	n/a
Z_dWS	90th - 10th	447.0000	n/a	419.0000	n/a	419.0000	n/a
Z_dWS	75th - 25th	80.0000	n/a	75.0000	n/a	76.0000	n/a

Table F.16: Distribution of Bias Measures for County by 2-Digit SIC: M and B

Relevant Variable	Statistic	Raw No Establishment Weight Applied Unweighted Bias	Weighted Bias	Raw Establishment Weight Applied Unweighted Bias	Weighted Bias	Fuzzed Establishment Weight Applied Unweighted Bias	Weighted Bias
M	Number of Obs	2156566		2143406		2142561	
M	Mean	1.3465	-0.0378	0.0949	-0.0086	0.0902	-0.0088
M	Variance	1095.2980	32.1823	1.6102	0.1097	1.6644	0.1142
M	Number of Missing Obs	896378		914994		915839	
M	99th Percentile	16.0000	0.9107	2.4000	0.7500	2.3333	0.7477
M	95th Percentile	1.3333	0.2578	0.6667	0.2727	0.6667	0.2734
M	90th Percentile	0.4444	0.1408	0.3043	0.1543	0.2857	0.1538
M	75th Percentile	0.0093	0.0301	0.0000	0.0356	0.0000	0.0368
M	50th Percentile	0.0000	-0.0077	0.0000	-0.0074	0.0000	-0.0070
M	25th Percentile	-0.0385	-0.0921	0.0000	-0.0571	0.0000	-0.0588
M	10th Percentile	-0.6667	-0.8811	-0.1875	-0.1667	-0.2000	-0.1692
M	5th Percentile	-1.0000	-0.9976	-0.4000	-0.3364	-0.4167	-0.3404
M	1st Percentile	-1.0000	-1.0000	-1.0000	-0.8753	-1.0000	-0.8650
M	90th - 10th	1.1111	1.0219	0.4918	0.3210	0.4857	0.3231
M	75th - 25th	0.0477	0.1223	0.0000	0.0927	0.0000	0.0957
B	Number of Obs	1997706		1985207		1984670	
B	Mean	1.2350	-0.0362	0.0892	-0.0017	0.0842	-0.0018
B	Variance	946.3363	31.7020	1.3247	0.1075	1.3759	0.1116
B	Number of Missing Obs	1055238		1073193		1073730	
B	99th Percentile	15.0000	0.9466	2.2500	0.7628	2.1667	0.7654
B	95th Percentile	1.2143	0.2619	0.6667	0.2798	0.6667	0.2798
B	90th Percentile	0.4211	0.1447	0.3153	0.1600	0.2941	0.1594
B	75th Percentile	0.0000	0.0386	0.0000	0.0435	0.0000	0.0448
B	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
B	25th Percentile	-0.0392	-0.0894	0.0000	-0.0476	0.0000	-0.0492
B	10th Percentile	-0.6786	-0.9307	-0.2000	-0.1618	-0.2000	-0.1642
B	5th Percentile	-1.0000	-0.9985	-0.4118	-0.3333	-0.4330	-0.3333
B	1st Percentile	-1.0000	-1.0000	-1.0000	-0.8938	-1.0000	-0.8797
B	90th - 10th	1.0996	1.0755	0.5153	0.3218	0.4941	0.3236
B	75th - 25th	0.0392	0.1280	0.0000	0.0911	0.0000	0.0939

Table F.17: Distribution of Bias Measures for County by 2-Digit SIC: E and F

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
E	Number of Obs	2001019		1988642		1988144	
E	Mean	1.2181	-0.0366	0.0879	-0.0026	0.0827	-0.0028
E	Variance	941.8433	31.6664	1.4195	0.1089	1.4480	0.1126
E	Number of Missing Obs	1051925		1069758		1070256	
E	99th Percentile	14.5000	0.9348	2.2000	0.7600	2.1429	0.7616
E	95th Percentile	1.1667	0.2614	0.6667	0.2768	0.6667	0.2778
E	90th Percentile	0.4000	0.1429	0.3077	0.1588	0.2889	0.1581
E	75th Percentile	0.0000	0.0371	0.0000	0.0412	0.0000	0.0424
E	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
E	25th Percentile	-0.0400	-0.0889	0.0000	-0.0479	0.0000	-0.0492
E	10th Percentile	-0.6667	-0.9237	-0.2000	-0.1623	-0.2000	-0.1655
E	5th Percentile	-1.0000	-0.9984	-0.4182	-0.3333	-0.4416	-0.3347
E	1st Percentile	-1.0000	-1.0000	-1.0000	-0.8933	-1.0000	-0.8795
E	90th - 10th	1.0667	1.0666	0.5077	0.3211	0.4889	0.3236
E	75th - 25th	0.0400	0.1260	0.0000	0.0891	0.0000	0.0917
F	Number of Obs	1861570		1850724		1850201	
F	Mean	1.1460	-0.0378	0.0850	0.0009	0.0799	0.0007
F	Variance	871.8582	31.8671	1.3869	0.1159	1.4501	0.1208
F	Number of Missing Obs	1191374		1207676		1208199	
F	99th Percentile	13.7500	0.9828	2.0000	0.7660	2.0000	0.7692
F	95th Percentile	1.0000	0.2627	0.6667	0.2809	0.6667	0.2814
F	90th Percentile	0.4000	0.1474	0.3143	0.1629	0.2941	0.1621
F	75th Percentile	0.0000	0.0428	0.0000	0.0476	0.0000	0.0488
F	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F	25th Percentile	-0.0411	-0.0903	0.0000	-0.0448	0.0000	-0.0463
F	10th Percentile	-0.7273	-0.9531	-0.2000	-0.1600	-0.2000	-0.1628
F	5th Percentile	-1.0000	-0.9989	-0.4348	-0.3333	-0.4667	-0.3333
F	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9103	-1.0000	-0.8919
F	90th - 10th	1.1273	1.1005	0.5143	0.3229	0.4941	0.3249
F	75th - 25th	0.0411	0.1331	0.0000	0.0924	0.0000	0.0951

Table F.18: Distribution of Bias Measures for County by 2-Digit SIC: W1 and Z_NA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
W1	Number of Obs	2167319		2150830		2150830	
W1	Mean	10.6523	0.2122	0.5101	0.0120	0.5185	0.0125
W1	Variance	1268862.1179	26767.0924	830.4980	16.6898	897.7869	18.0849
W1	Number of Missing Obs	885625		907570		907570	
W1	99th Percentile	42.7371	1.1677	5.7010	0.9415	5.6735	0.9504
W1	95th Percentile	2.5794	0.3072	0.9245	0.3247	0.9224	0.3267
W1	90th Percentile	0.6429	0.1578	0.3780	0.1739	0.3771	0.1737
W1	75th Percentile	0.0684	0.0338	0.0473	0.0389	0.0473	0.0398
W1	50th Percentile	0.0000	-0.0068	-0.0068	-0.0086	-0.0068	-0.0086
W1	25th Percentile	-0.0489	-0.0920	-0.0106	-0.0537	-0.0111	-0.0549
W1	10th Percentile	-0.6956	-0.8969	-0.2040	-0.1720	-0.2045	-0.1733
W1	5th Percentile	-0.9947	-0.9986	-0.4409	-0.3486	-0.4412	-0.3486
W1	1st Percentile	-1.0000	-1.0000	-0.9172	-0.8972	-0.9175	-0.8905
W1	90th - 10th	1.3385	1.0547	0.5820	0.3458	0.5815	0.3470
W1	75th - 25th	0.1173	0.1258	0.0579	0.0925	0.0584	0.0947
Z_NA	Number of Obs	995988		990662		990662	
Z_NA	Mean	-0.0366	0.0020	0.0237	0.0364	0.0247	0.0360
Z_NA	Variance	0.3873	0.4087	0.1896	0.2397	0.1911	0.2104
Z_NA	Number of Missing Obs	2056956		2067738		2067738	
Z_NA	99th Percentile	1.9500	1.3947	1.3810	0.8421	1.4071	0.8669
Z_NA	95th Percentile	0.5071	0.3008	0.4286	0.2453	0.4349	0.2533
Z_NA	90th Percentile	0.2174	0.1353	0.1927	0.1266	0.1966	0.1304
Z_NA	75th Percentile	0.0063	0.0438	0.0104	0.0449	0.0132	0.0442
Z_NA	50th Percentile	0.0000	0.0038	0.0000	0.0066	0.0000	0.0056
Z_NA	25th Percentile	-0.0746	-0.0280	-0.0192	-0.0195	-0.0208	-0.0196
Z_NA	10th Percentile	-0.7160	-0.1500	-0.1714	-0.0739	-0.1748	-0.0783
Z_NA	5th Percentile	-1.0000	-1.0000	-0.3448	-0.1364	-0.3488	-0.1423
Z_NA	1st Percentile	-1.0000	-1.0000	-1.0000	-0.4000	-1.0000	-0.4104
Z_NA	90th - 10th	0.9334	0.2853	0.3641	0.2005	0.3714	0.2087
Z_NA	75th - 25th	0.0809	0.0718	0.0296	0.0644	0.0341	0.0638

Table F.19: Distribution of Bias Measures for County by 2-Digit SIC: Z_NH and Z_NR

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NH	Number of Obs	875806		871165		871165	
Z_NH	Mean	-0.0430	0.0041	0.0211	0.0421	0.0223	0.0418
Z_NH	Variance	0.4027	0.4590	0.1879	0.1882	0.1953	0.1860
Z_NH	Number of Missing Obs	2177138		2187235		2187235	
Z_NH	99th Percentile	1.9000	1.3810	1.4074	0.9275	1.4362	0.9356
Z_NH	95th Percentile	0.5038	0.3148	0.4370	0.2667	0.4455	0.2738
Z_NH	90th Percentile	0.2150	0.1453	0.1971	0.1400	0.2006	0.1435
Z_NH	75th Percentile	0.0043	0.0513	0.0105	0.0533	0.0141	0.0521
Z_NH	50th Percentile	0.0000	0.0071	0.0000	0.0093	0.0000	0.0096
Z_NH	25th Percentile	-0.0846	-0.0275	-0.0245	-0.0181	-0.0262	-0.0187
Z_NH	10th Percentile	-0.8325	-0.1576	-0.1875	-0.0765	-0.1906	-0.0812
Z_NH	5th Percentile	-1.0000	-1.0000	-0.3750	-0.1418	-0.3779	-0.1481
Z_NH	1st Percentile	-1.0000	-1.0000	-1.0000	-0.4367	-1.0000	-0.4459
Z_NH	90th - 10th	1.0475	0.3030	0.3846	0.2165	0.3911	0.2247
Z_NH	75th - 25th	0.0890	0.0788	0.0350	0.0714	0.0404	0.0709
Z_NR	Number of Obs	449712		448407		448407	
Z_NR	Mean	-0.0971	-0.0436	-0.0123	0.0116	-0.0117	0.0118
Z_NR	Variance	0.2124	0.3554	0.0830	0.0755	0.0846	0.0768
Z_NR	Number of Missing Obs	2603232		2609993		2609993	
Z_NR	99th Percentile	1.0000	1.0750	0.9556	0.8644	0.9565	0.8785
Z_NR	95th Percentile	0.2727	0.2500	0.2600	0.2333	0.2620	0.2358
Z_NR	90th Percentile	0.0800	0.1081	0.0850	0.1081	0.0885	0.1107
Z_NR	75th Percentile	0.0000	0.0189	0.0000	0.0211	0.0000	0.0210
Z_NR	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_NR	25th Percentile	-0.0526	-0.0427	0.0000	-0.0235	-0.0027	-0.0239
Z_NR	10th Percentile	-1.0000	-0.2900	-0.1517	-0.1000	-0.1542	-0.1009
Z_NR	5th Percentile	-1.0000	-1.0000	-0.3567	-0.1915	-0.3605	-0.1955
Z_NR	1st Percentile	-1.0000	-1.0000	-1.0000	-0.8372	-1.0000	-0.8943
Z_NR	90th - 10th	1.0800	0.3981	0.2367	0.2081	0.2427	0.2115
Z_NR	75th - 25th	0.0526	0.0616	0.0000	0.0446	0.0027	0.0449

Table F.20: Distribution of Bias Measures for County by 2-Digit SIC: Z_NS and Z_W2

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NS	Number of Obs	1005002		1000214		1000214	
Z_NS	Mean	-0.0358	0.0018	0.0213	0.0360	0.0223	0.0359
Z_NS	Variance	0.3801	0.2338	0.1545	0.0857	0.1613	0.0892
Z_NS	Number of Missing Obs	2047942		2058186		2058186	
Z_NS	99th Percentile	1.9400	1.3952	1.3789	0.8163	1.4048	0.8453
Z_NS	95th Percentile	0.5075	0.2955	0.4273	0.2444	0.4323	0.2478
Z_NS	90th Percentile	0.2143	0.1377	0.1908	0.1287	0.1940	0.1310
Z_NS	75th Percentile	0.0039	0.0526	0.0086	0.0530	0.0120	0.0510
Z_NS	50th Percentile	0.0000	0.0079	0.0000	0.0092	0.0000	0.0099
Z_NS	25th Percentile	-0.0750	-0.0252	-0.0191	-0.0163	-0.0209	-0.0170
Z_NS	10th Percentile	-0.6850	-0.1576	-0.1780	-0.0723	-0.1809	-0.0756
Z_NS	5th Percentile	-1.0000	-1.0000	-0.3548	-0.1391	-0.3596	-0.1436
Z_NS	1st Percentile	-1.0000	-1.0000	-1.0000	-0.4194	-1.0000	-0.4277
Z_NS	90th - 10th	0.8993	0.2953	0.3688	0.2010	0.3749	0.2066
Z_NS	75th - 25th	0.0789	0.0778	0.0277	0.0692	0.0329	0.0679
Z_W2	Number of Obs	2008038		1993843		1993843	
Z_W2	Mean	0.1908	-0.0374	0.1107	0.0061	0.1129	0.0056
Z_W2	Variance	68.5005	2.7928	11.1844	0.3178	14.0305	0.3829
Z_W2	Number of Missing Obs	1044906		1064557		1064557	
Z_W2	99th Percentile	3.8031	0.5571	1.8354	0.3705	1.8571	0.3797
Z_W2	95th Percentile	0.6382	0.1388	0.3422	0.1207	0.3461	0.1265
Z_W2	90th Percentile	0.2316	0.0660	0.1442	0.0632	0.1463	0.0660
Z_W2	75th Percentile	0.0211	0.0137	0.0134	0.0156	0.0141	0.0161
Z_W2	50th Percentile	0.0000	-0.0033	0.0000	-0.0007	0.0000	-0.0014
Z_W2	25th Percentile	-0.0285	-0.0324	-0.0052	-0.0204	-0.0071	-0.0218
Z_W2	10th Percentile	-0.2986	-0.2104	-0.0928	-0.0594	-0.0956	-0.0631
Z_W2	5th Percentile	-0.9472	-0.5441	-0.1994	-0.1085	-0.2032	-0.1153
Z_W2	1st Percentile	-1.0000	-1.0000	-0.6246	-0.2857	-0.6272	-0.3023
Z_W2	90th - 10th	0.5301	0.2764	0.2369	0.1226	0.2419	0.1291
Z_W2	75th - 25th	0.0497	0.0461	0.0186	0.0360	0.0212	0.0379

Table F.21: Distribution of Bias Measures for County by 2-Digit SIC: Z_W3 and Z_WFA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_W3	Number of Obs	1868665		1856094		1856094	
Z_W3	Mean	0.1588	-0.0450	0.0978	0.0026	0.0998	0.0021
Z_W3	Variance	50.8150	2.4371	11.8337	0.3685	14.8942	0.4450
Z_W3	Number of Missing Obs	1184279		1202306		1202306	
Z_W3	99th Percentile	3.2338	0.5296	1.6226	0.3524	1.6452	0.3612
Z_W3	95th Percentile	0.5771	0.1300	0.3185	0.1120	0.3224	0.1188
Z_W3	90th Percentile	0.2139	0.0603	0.1356	0.0578	0.1379	0.0609
Z_W3	75th Percentile	0.0187	0.0108	0.0120	0.0126	0.0129	0.0134
Z_W3	50th Percentile	0.0000	-0.0052	0.0000	-0.0025	0.0000	-0.0029
Z_W3	25th Percentile	-0.0278	-0.0374	-0.0050	-0.0228	-0.0068	-0.0243
Z_W3	10th Percentile	-0.2957	-0.2360	-0.0873	-0.0621	-0.0903	-0.0659
Z_W3	5th Percentile	-1.0000	-0.5752	-0.1879	-0.1109	-0.1923	-0.1181
Z_W3	1st Percentile	-1.0000	-1.0000	-0.6104	-0.2813	-0.6127	-0.3012
Z_W3	90th - 10th	0.5096	0.2964	0.2229	0.1198	0.2283	0.1268
Z_W3	75th - 25th	0.0464	0.0482	0.0171	0.0354	0.0197	0.0377
Z_WFA	Number of Obs	878424		874403		874403	
Z_WFA	Mean	0.1348	-0.0569	0.1655	0.0041	0.1696	0.0049
Z_WFA	Variance	46.2491	5.2637	32.7026	3.4760	36.3225	3.8860
Z_WFA	Number of Missing Obs	2174520		2183997		2183997	
Z_WFA	99th Percentile	3.7339	0.9395	2.9330	0.7455	2.9858	0.7719
Z_WFA	95th Percentile	0.6397	0.2086	0.5276	0.1962	0.5376	0.2032
Z_WFA	90th Percentile	0.2332	0.0901	0.2070	0.0917	0.2113	0.0948
Z_WFA	75th Percentile	0.0032	0.0084	0.0064	0.0124	0.0092	0.0132
Z_WFA	50th Percentile	0.0000	-0.0179	0.0000	-0.0108	0.0000	-0.0115
Z_WFA	25th Percentile	-0.0649	-0.1040	-0.0135	-0.0651	-0.0154	-0.0673
Z_WFA	10th Percentile	-0.8002	-0.3634	-0.1517	-0.1566	-0.1543	-0.1587
Z_WFA	5th Percentile	-1.0000	-1.0000	-0.3088	-0.2247	-0.3124	-0.2267
Z_WFA	1st Percentile	-1.0000	-1.0000	-1.0000	-0.5287	-1.0000	-0.5365
Z_WFA	90th - 10th	1.0334	0.4535	0.3587	0.2483	0.3656	0.2535
Z_WFA	75th - 25th	0.0681	0.1124	0.0199	0.0775	0.0246	0.0805

Table F.22: Distribution of Bias Measures for County by 2-Digit SIC: Z_WFS and Z_WH3

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_WFS	Number of Obs	835013		831524		831524	
Z_WFS	Mean	0.5259	0.0441	0.4840	0.0925	0.4945	0.0939
Z_WFS	Variance	204.2645	28.3196	152.0791	19.8959	197.9968	24.8947
Z_WFS	Number of Missing Obs	2217931		2226876		2226876	
Z_WFS	99th Percentile	9.5915	1.7831	7.5263	1.4527	7.5474	1.4655
Z_WFS	95th Percentile	1.1852	0.3636	0.9745	0.3512	0.9812	0.3565
Z_WFS	90th Percentile	0.3777	0.1655	0.3433	0.1714	0.3451	0.1721
Z_WFS	75th Percentile	0.0046	0.0432	0.0119	0.0491	0.0145	0.0483
Z_WFS	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_WFS	25th Percentile	-0.0704	-0.0660	-0.0038	-0.0394	-0.0081	-0.0404
Z_WFS	10th Percentile	-1.0000	-0.4339	-0.1782	-0.1333	-0.1796	-0.1355
Z_WFS	5th Percentile	-1.0000	-1.0000	-0.3995	-0.2425	-0.4029	-0.2474
Z_WFS	1st Percentile	-1.0000	-1.0000	-1.0000	-0.6180	-1.0000	-0.6222
Z_WFS	90th - 10th	1.3777	0.5995	0.5215	0.3047	0.5247	0.3076
Z_WFS	75th - 25th	0.0749	0.1092	0.0157	0.0885	0.0227	0.0887
Z_WH3	Number of Obs	728111		724717		724717	
Z_WH3	Mean	0.0882	-0.0702	0.1321	-0.0046	0.1364	-0.0040
Z_WH3	Variance	27.4092	3.3557	14.3059	1.6333	15.6135	1.8288
Z_WH3	Number of Missing Obs	2324833		2333683		2333683	
Z_WH3	99th Percentile	3.3125	0.8880	2.6762	0.7195	2.7462	0.7545
Z_WH3	95th Percentile	0.6161	0.2117	0.5225	0.2026	0.5331	0.2095
Z_WH3	90th Percentile	0.2306	0.0930	0.2092	0.0954	0.2146	0.0986
Z_WH3	75th Percentile	0.0040	0.0085	0.0082	0.0127	0.0109	0.0136
Z_WH3	50th Percentile	0.0000	-0.0224	0.0000	-0.0139	0.0000	-0.0144
Z_WH3	25th Percentile	-0.0758	-0.1171	-0.0184	-0.0726	-0.0201	-0.0751
Z_WH3	10th Percentile	-1.0000	-0.4020	-0.1618	-0.1699	-0.1646	-0.1717
Z_WH3	5th Percentile	-1.0000	-1.0000	-0.3236	-0.2406	-0.3273	-0.2446
Z_WH3	1st Percentile	-1.0000	-1.0000	-1.0000	-0.5634	-1.0000	-0.5565
Z_WH3	90th - 10th	1.2306	0.4950	0.3711	0.2653	0.3791	0.2703
Z_WH3	75th - 25th	0.0799	0.1256	0.0266	0.0853	0.0310	0.0887

Table F.23: Distribution of Bias Measures for County by 2-Digit SIC: A and FA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
A	Number of Obs	2978784	n/a	2984208	n/a	2984208	n/a
A	Mean	-0.2261	n/a	-0.2574	n/a	-0.2660	n/a
A	Variance	299.4617	n/a	166.0879	n/a	170.3244	n/a
A	Number of Missing Obs	74160	n/a	74192	n/a	74192	n/a
A	99th Percentile	15.0000	n/a	9.0000	n/a	9.0000	n/a
A	95th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
A	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
A	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	10th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
A	5th Percentile	-2.0000	n/a	-1.0000	n/a	-1.0000	n/a
A	1st Percentile	-17.0000	n/a	-11.0000	n/a	-11.0000	n/a
A	90th - 10th	2.0000	n/a	1.0000	n/a	1.0000	n/a
A	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	Number of Obs	2830704	n/a	2835888	n/a	2835888	n/a
FA	Mean	0.1599	n/a	0.1641	n/a	0.1616	n/a
FA	Variance	118.0914	n/a	61.3042	n/a	63.0941	n/a
FA	Number of Missing Obs	222240	n/a	222512	n/a	222512	n/a
FA	99th Percentile	12.0000	n/a	8.0000	n/a	8.0000	n/a
FA	95th Percentile	2.0000	n/a	1.0000	n/a	1.0000	n/a
FA	90th Percentile	1.0000	n/a	0.0000	n/a	0.0000	n/a
FA	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FA	1st Percentile	-8.0000	n/a	-4.0000	n/a	-4.0000	n/a
FA	90th - 10th	1.0000	n/a	0.0000	n/a	0.0000	n/a
FA	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.24: Distribution of Bias Measures for County by 2-Digit SIC: FJC and FJD

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJC	Number of Obs	2830704	n/a	2835888	n/a	1989970	n/a
FJC	Mean	0.1000	n/a	0.1005	n/a	0.1412	n/a
FJC	Variance	92.5667	n/a	57.8056	n/a	82.0907	n/a
FJC	Number of Missing Obs	222240	n/a	222512	n/a	1068430	n/a
FJC	99th Percentile	9.0000	n/a	6.0000	n/a	8.0000	n/a
FJC	95th Percentile	1.0000	n/a	1.0000	n/a	2.0000	n/a
FJC	90th Percentile	0.0000	n/a	0.0000	n/a	1.0000	n/a
FJC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FJC	1st Percentile	-6.0000	n/a	-4.0000	n/a	-5.0000	n/a
FJC	90th - 10th	0.0000	n/a	0.0000	n/a	1.0000	n/a
FJC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	Number of Obs	2830704	n/a	2835888	n/a	1989970	n/a
FJD	Mean	-0.1473	n/a	-0.1515	n/a	-0.2200	n/a
FJD	Variance	35.0832	n/a	17.8442	n/a	25.8026	n/a
FJD	Number of Missing Obs	222240	n/a	222512	n/a	1068430	n/a
FJD	99th Percentile	4.0000	n/a	2.0000	n/a	3.0000	n/a
FJD	95th Percentile	1.0000	n/a	0.0000	n/a	1.0000	n/a
FJD	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FJD	1st Percentile	-7.0000	n/a	-5.0000	n/a	-7.0000	n/a
FJD	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.25: Distribution of Bias Measures for County by 2-Digit SIC: FJF and FS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJF	Number of Obs	2830704	n/a	2835888	n/a	1989970	n/a
FJF	Mean	0.2462	n/a	0.2511	n/a	0.3599	n/a
FJF	Variance	113.1506	n/a	86.3410	n/a	123.1210	n/a
FJF	Number of Missing Obs	222240	n/a	222512	n/a	1068430	n/a
FJF	99th Percentile	11.0000	n/a	9.0000	n/a	13.0000	n/a
FJF	95th Percentile	2.0000	n/a	2.0000	n/a	2.0000	n/a
FJF	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FJF	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	10th Percentile	0.0000	n/a	0.0000	n/a	-1.0000	n/a
FJF	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FJF	1st Percentile	-7.0000	n/a	-4.0000	n/a	-6.0000	n/a
FJF	90th - 10th	1.0000	n/a	1.0000	n/a	2.0000	n/a
FJF	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	Number of Obs	2830704	n/a	2835888	n/a	2835888	n/a
FS	Mean	-0.0881	n/a	-0.0880	n/a	-0.0927	n/a
FS	Variance	52.6652	n/a	20.6787	n/a	21.2441	n/a
FS	Number of Missing Obs	222240	n/a	222512	n/a	222512	n/a
FS	99th Percentile	6.0000	n/a	3.0000	n/a	3.0000	n/a
FS	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FS	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
FS	1st Percentile	-8.0000	n/a	-5.0000	n/a	-5.0000	n/a
FS	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.26: Distribution of Bias Measures for County by 2-Digit SIC: FT and H

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FT	Number of Obs	2796688	n/a	2785760	n/a	2785760	n/a
FT	Mean	0.0158	n/a	0.0172	n/a	0.0174	n/a
FT	Variance	0.9278	n/a	0.6656	n/a	0.6864	n/a
FT	Number of Missing Obs	256256	n/a	272640	n/a	272640	n/a
FT	99th Percentile	0.5000	n/a	0.5000	n/a	0.5000	n/a
FT	95th Percentile	0.1900	n/a	0.1000	n/a	0.0990	n/a
FT	90th Percentile	0.0700	n/a	0.0400	n/a	0.0360	n/a
FT	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	10th Percentile	-0.0500	n/a	-0.0200	n/a	-0.0160	n/a
FT	5th Percentile	-0.1600	n/a	-0.0600	n/a	-0.0630	n/a
FT	1st Percentile	-0.5000	n/a	-0.3500	n/a	-0.3470	n/a
FT	90th - 10th	0.1200	n/a	0.0600	n/a	0.0520	n/a
FT	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	Number of Obs	2756256	n/a	2761520	n/a	2761520	n/a
H	Mean	-0.2302	n/a	-0.2564	n/a	-0.2647	n/a
H	Variance	233.2889	n/a	147.6522	n/a	151.8700	n/a
H	Number of Missing Obs	296688	n/a	296880	n/a	296880	n/a
H	99th Percentile	13.0000	n/a	8.0000	n/a	8.0000	n/a
H	95th Percentile	2.0000	n/a	1.0000	n/a	1.0000	n/a
H	90th Percentile	1.0000	n/a	1.0000	n/a	0.0000	n/a
H	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	10th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
H	5th Percentile	-2.0000	n/a	-1.0000	n/a	-1.0000	n/a
H	1st Percentile	-15.0000	n/a	-10.0000	n/a	-10.0000	n/a
H	90th - 10th	2.0000	n/a	1.0000	n/a	0.0000	n/a
H	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.27: Distribution of Bias Measures for County by 2-Digit SIC: H3 and JC

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
H3	Number of Obs	2608048	n/a	2613104	n/a	2613104	n/a
H3	Mean	0.1568	n/a	0.1639	n/a	0.1614	n/a
H3	Variance	96.0231	n/a	56.5862	n/a	58.3745	n/a
H3	Number of Missing Obs	444896	n/a	445296	n/a	445296	n/a
H3	99th Percentile	10.0000	n/a	7.0000	n/a	7.0000	n/a
H3	95th Percentile	2.0000	n/a	1.0000	n/a	1.0000	n/a
H3	90th Percentile	1.0000	n/a	0.0000	n/a	0.0000	n/a
H3	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
H3	1st Percentile	-7.0000	n/a	-4.0000	n/a	-4.0000	n/a
H3	90th - 10th	1.0000	n/a	0.0000	n/a	0.0000	n/a
H3	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	Number of Obs	2905040	n/a	2910304	n/a	2137865	n/a
JC	Mean	-0.1291	n/a	-0.1388	n/a	-0.1943	n/a
JC	Variance	50.7931	n/a	40.6362	n/a	56.3760	n/a
JC	Number of Missing Obs	147904	n/a	148096	n/a	920535	n/a
JC	99th Percentile	6.0000	n/a	4.0000	n/a	5.0000	n/a
JC	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JC	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	10th Percentile	0.0000	n/a	0.0000	n/a	-1.0000	n/a
JC	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
JC	1st Percentile	-9.0000	n/a	-6.0000	n/a	-8.0000	n/a
JC	90th - 10th	0.0000	n/a	0.0000	n/a	1.0000	n/a
JC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.28: Distribution of Bias Measures for County by 2-Digit SIC: JD and JF

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
JD	Number of Obs	2905040	n/a	2910304	n/a	2137865	n/a
JD	Mean	-0.1051	n/a	-0.1104	n/a	-0.1548	n/a
JD	Variance	52.3663	n/a	23.8528	n/a	32.6098	n/a
JD	Number of Missing Obs	147904	n/a	148096	n/a	920535	n/a
JD	99th Percentile	6.0000	n/a	3.0000	n/a	4.0000	n/a
JD	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JD	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	10th Percentile	0.0000	n/a	0.0000	n/a	-1.0000	n/a
JD	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
JD	1st Percentile	-8.0000	n/a	-5.0000	n/a	-7.0000	n/a
JD	90th - 10th	0.0000	n/a	0.0000	n/a	1.0000	n/a
JD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	Number of Obs	2905040	n/a	2910304	n/a	2137865	n/a
JF	Mean	-0.0237	n/a	-0.0279	n/a	-0.0391	n/a
JF	Variance	71.4032	n/a	58.0607	n/a	80.2101	n/a
JF	Number of Missing Obs	147904	n/a	148096	n/a	920535	n/a
JF	99th Percentile	10.0000	n/a	7.0000	n/a	9.0000	n/a
JF	95th Percentile	2.0000	n/a	1.0000	n/a	2.0000	n/a
JF	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JF	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	10th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
JF	5th Percentile	-2.0000	n/a	-1.0000	n/a	-2.0000	n/a
JF	1st Percentile	-10.0000	n/a	-7.0000	n/a	-10.0000	n/a
JF	90th - 10th	2.0000	n/a	1.0000	n/a	2.0000	n/a
JF	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.29: Distribution of Bias Measures for County by 2-Digit SIC: R and S

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
R	Number of Obs	2756256	n/a	2761520	n/a	2761520	n/a
R	Mean	-0.0144	n/a	-0.0115	n/a	-0.0152	n/a
R	Variance	13.5232	n/a	8.6704	n/a	8.6151	n/a
R	Number of Missing Obs	296688	n/a	296880	n/a	296880	n/a
R	99th Percentile	3.0000	n/a	2.0000	n/a	2.0000	n/a
R	95th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	5th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	1st Percentile	-3.0000	n/a	-2.0000	n/a	-2.0000	n/a
R	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	Number of Obs	2979200	n/a	2984496	n/a	2984496	n/a
S	Mean	-0.2020	n/a	-0.2281	n/a	-0.2345	n/a
S	Variance	278.5617	n/a	126.4811	n/a	129.4682	n/a
S	Number of Missing Obs	73744	n/a	73904	n/a	73904	n/a
S	99th Percentile	15.0000	n/a	8.0000	n/a	8.0000	n/a
S	95th Percentile	2.0000	n/a	1.0000	n/a	1.0000	n/a
S	90th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
S	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	10th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
S	5th Percentile	-2.0000	n/a	-1.0000	n/a	-1.0000	n/a
S	1st Percentile	-17.0000	n/a	-10.0000	n/a	-10.0000	n/a
S	90th - 10th	2.0000	n/a	1.0000	n/a	1.0000	n/a
S	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.30: Distribution of Bias Measures for County by 2-Digit SIC: Z_dWA and Z_dWS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_dWA	Number of Obs	1545454	n/a	1455552	n/a	1455552	n/a
Z_dWA	Mean	104.4956	n/a	112.5215	n/a	116.0884	n/a
Z_dWA	Variance	618891796.3897	n/a	634134444.2969	n/a	793196069.3331	n/a
Z_dWA	Number of Missing Obs	1507490	n/a	1602848	n/a	1602848	n/a
Z_dWA	99th Percentile	3292.0000	n/a	2676.0000	n/a	2681.0000	n/a
Z_dWA	95th Percentile	1036.0000	n/a	765.0000	n/a	764.0000	n/a
Z_dWA	90th Percentile	520.0000	n/a	353.0000	n/a	351.0000	n/a
Z_dWA	75th Percentile	91.0000	n/a	47.0000	n/a	47.0000	n/a
Z_dWA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWA	25th Percentile	-19.0000	n/a	-1.0000	n/a	-3.0000	n/a
Z_dWA	10th Percentile	-300.0000	n/a	-141.0000	n/a	-140.0000	n/a
Z_dWA	5th Percentile	-702.0000	n/a	-363.0000	n/a	-363.0000	n/a
Z_dWA	1st Percentile	-2494.0000	n/a	-1524.0000	n/a	-1530.0000	n/a
Z_dWA	90th - 10th	820.0000	n/a	494.0000	n/a	491.0000	n/a
Z_dWA	75th - 25th	110.0000	n/a	48.0000	n/a	50.0000	n/a
Z_dWS	Number of Obs	1553676	n/a	1463051	n/a	1463051	n/a
Z_dWS	Mean	-107.6754	n/a	-117.9538	n/a	-121.7419	n/a
Z_dWS	Variance	220630242.6013	n/a	259593368.0573	n/a	315358191.9986	n/a
Z_dWS	Number of Missing Obs	1499268	n/a	1595349	n/a	1595349	n/a
Z_dWS	99th Percentile	2597.0000	n/a	1574.0000	n/a	1573.0000	n/a
Z_dWS	95th Percentile	744.0000	n/a	394.0000	n/a	394.0000	n/a
Z_dWS	90th Percentile	318.0000	n/a	158.0000	n/a	158.0000	n/a
Z_dWS	75th Percentile	24.0000	n/a	3.0000	n/a	5.0000	n/a
Z_dWS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWS	25th Percentile	-82.0000	n/a	-42.0000	n/a	-42.0000	n/a
Z_dWS	10th Percentile	-523.0000	n/a	-339.0000	n/a	-338.0000	n/a
Z_dWS	5th Percentile	-1092.0000	n/a	-781.0000	n/a	-781.0000	n/a
Z_dWS	1st Percentile	-3584.0000	n/a	-2922.0000	n/a	-2946.0000	n/a
Z_dWS	90th - 10th	841.0000	n/a	497.0000	n/a	496.0000	n/a
Z_dWS	75th - 25th	106.0000	n/a	45.0000	n/a	47.0000	n/a

Table F.31: Distribution of Bias Measures for County by 3-Digit SIC: M and B

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
M	Number of Obs	4659205		4567707		4565036	
M	Mean	1.3293	-0.1062	0.0884	-0.0144	0.0841	-0.0149
M	Variance	898.2522	60.2392	2.1364	0.2389	2.2335	0.2502
M	Number of Missing Obs	3905883		4084853		4087524	
M	99th Percentile	21.0000	2.1667	2.7500	0.9850	2.7027	0.9846
M	95th Percentile	2.0000	0.2955	0.6923	0.3044	0.6667	0.3014
M	90th Percentile	0.5000	0.1250	0.2500	0.1615	0.2222	0.1587
M	75th Percentile	0.0000	0.0074	0.0000	0.0313	0.0000	0.0308
M	50th Percentile	0.0000	-0.0413	0.0000	-0.0038	0.0000	-0.0040
M	25th Percentile	-0.2450	-0.9326	0.0000	-0.0730	0.0000	-0.0755
M	10th Percentile	-1.0000	-0.9995	-0.2000	-0.2314	-0.2308	-0.2321
M	5th Percentile	-1.0000	-1.0000	-0.5000	-0.4744	-0.5000	-0.4753
M	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9738	-1.0000	-0.9736
M	90th - 10th	1.5000	1.1245	0.4500	0.3929	0.4530	0.3908
M	75th - 25th	0.2450	0.9399	0.0000	0.1044	0.0000	0.1062
B	Number of Obs	4214874		4127836		4126178	
B	Mean	1.1735	-0.1089	0.0819	-0.0077	0.0774	-0.0082
B	Variance	636.7395	49.3363	1.8494	0.2410	1.9402	0.2532
B	Number of Missing Obs	4350214		4524724		4526382	
B	99th Percentile	19.0000	2.1667	2.5000	1.0000	2.5000	1.0000
B	95th Percentile	2.0000	0.3043	0.6842	0.3146	0.6667	0.3121
B	90th Percentile	0.5000	0.1284	0.2500	0.1685	0.2308	0.1667
B	75th Percentile	0.0000	0.0128	0.0000	0.0399	0.0000	0.0396
B	50th Percentile	0.0000	-0.0351	0.0000	0.0000	0.0000	0.0000
B	25th Percentile	-0.2667	-0.9383	0.0000	-0.0628	0.0000	-0.0639
B	10th Percentile	-1.0000	-1.0000	-0.2105	-0.2242	-0.2500	-0.2260
B	5th Percentile	-1.0000	-1.0000	-0.5000	-0.4740	-0.5000	-0.4744
B	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9750	-1.0000	-0.9748
B	90th - 10th	1.5000	1.1284	0.4605	0.3927	0.4808	0.3927
B	75th - 25th	0.2667	0.9510	0.0000	0.1027	0.0000	0.1035

Table F.32: Distribution of Bias Measures for County by 3-Digit SIC: E and F

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
E	Number of Obs	4222032		4136783		4135244	
E	Mean	1.1658	-0.1097	0.0801	-0.0087	0.0754	-0.0092
E	Variance	642.4059	49.4847	1.8293	0.2396	1.9013	0.2509
E	Number of Missing Obs	4343056		4515777		4517316	
E	99th Percentile	19.0000	2.1029	2.5000	1.0000	2.5000	1.0000
E	95th Percentile	2.0000	0.3000	0.6667	0.3120	0.6667	0.3085
E	90th Percentile	0.5000	0.1266	0.2500	0.1667	0.2222	0.1667
E	75th Percentile	0.0000	0.0122	0.0000	0.0388	0.0000	0.0385
E	50th Percentile	0.0000	-0.0344	0.0000	0.0000	0.0000	0.0000
E	25th Percentile	-0.2500	-0.9336	0.0000	-0.0635	0.0000	-0.0645
E	10th Percentile	-1.0000	-1.0000	-0.2143	-0.2241	-0.2500	-0.2265
E	5th Percentile	-1.0000	-1.0000	-0.5000	-0.4737	-0.5000	-0.4744
E	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9752	-1.0000	-0.9750
E	90th - 10th	1.5000	1.1266	0.4643	0.3908	0.4722	0.3932
E	75th - 25th	0.2500	0.9458	0.0000	0.1023	0.0000	0.1029
F	Number of Obs	3853664		3775370		3773813	
F	Mean	1.0660	-0.1142	0.0762	-0.0054	0.0714	-0.0061
F	Variance	537.3849	45.4892	1.8748	0.2571	1.9838	0.2715
F	Number of Missing Obs	4711424		4877190		4878747	
F	99th Percentile	17.3333	2.1250	2.5000	1.0000	2.4000	1.0000
F	95th Percentile	2.0000	0.3091	0.6667	0.3242	0.6667	0.3215
F	90th Percentile	0.5000	0.1290	0.2500	0.1765	0.2222	0.1739
F	75th Percentile	0.0000	0.0143	0.0000	0.0426	0.0000	0.0423
F	50th Percentile	0.0000	-0.0333	0.0000	0.0000	0.0000	0.0000
F	25th Percentile	-0.3333	-0.9432	0.0000	-0.0597	0.0000	-0.0608
F	10th Percentile	-1.0000	-1.0000	-0.2222	-0.2231	-0.2500	-0.2262
F	5th Percentile	-1.0000	-1.0000	-0.5000	-0.4773	-0.5000	-0.4772
F	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9759	-1.0000	-0.9761
F	90th - 10th	1.5000	1.1290	0.4722	0.3995	0.4722	0.4001
F	75th - 25th	0.3333	0.9575	0.0000	0.1023	0.0000	0.1030

Table F.33: Distribution of Bias Measures for County by 3-Digit SIC: W1 and Z_NA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
W1	Number of Obs	4692354		4590415		4590415	
W1	Mean	9.1409	0.3712	0.5425	0.0196	0.5521	0.0207
W1	Variance	478853.7551	22207.5617	1163.7311	41.7372	1235.3079	44.4022
W1	Number of Missing Obs	3872734		4062145		4062145	
W1	99th Percentile	60.6331	3.2674	6.2489	1.2237	6.3407	1.2351
W1	95th Percentile	4.2667	0.3653	0.9116	0.3636	0.9143	0.3637
W1	90th Percentile	0.8938	0.1513	0.3029	0.1830	0.3023	0.1819
W1	75th Percentile	0.0000	0.0113	0.0000	0.0325	0.0000	0.0324
W1	50th Percentile	0.0000	-0.0391	-0.0077	-0.0089	-0.0077	-0.0090
W1	25th Percentile	-0.2525	-0.9434	-0.0104	-0.0726	-0.0104	-0.0739
W1	10th Percentile	-1.0000	-0.9995	-0.2276	-0.2316	-0.2278	-0.2299
W1	5th Percentile	-1.0000	-1.0000	-0.5557	-0.4731	-0.5565	-0.4728
W1	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9728	-1.0000	-0.9723
W1	90th - 10th	1.8938	1.1508	0.5305	0.4146	0.5302	0.4118
W1	75th - 25th	0.2525	0.9546	0.0104	0.1051	0.0104	0.1062
Z_NA	Number of Obs	1653730		1625076		1625077	
Z_NA	Mean	-0.1360	-0.0653	0.0072	0.0370	0.0080	0.0368
Z_NA	Variance	0.5841	0.9894	0.1689	0.2239	0.1727	0.2296
Z_NA	Number of Missing Obs	6911358		7027484		7027483	
Z_NA	99th Percentile	2.3000	2.1746	1.3000	1.0294	1.3239	1.0465
Z_NA	95th Percentile	0.5488	0.6325	0.3636	0.2817	0.3700	0.2890
Z_NA	90th Percentile	0.1717	0.2663	0.1366	0.1429	0.1394	0.1466
Z_NA	75th Percentile	0.0000	0.0496	0.0000	0.0476	0.0000	0.0462
Z_NA	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_NA	25th Percentile	-0.3500	-0.1295	0.0000	-0.0205	-0.0018	-0.0210
Z_NA	10th Percentile	-1.0000	-1.0000	-0.1742	-0.0909	-0.1765	-0.0959
Z_NA	5th Percentile	-1.0000	-1.0000	-0.4033	-0.1777	-0.4064	-0.1854
Z_NA	1st Percentile	-1.0000	-1.0000	-1.0000	-0.6700	-1.0000	-0.6997
Z_NA	90th - 10th	1.1717	1.2663	0.3108	0.2338	0.3158	0.2424
Z_NA	75th - 25th	0.3500	0.1791	0.0000	0.0681	0.0018	0.0672

Table F.34: Distribution of Bias Measures for County by 3-Digit SIC: Z_NH and Z_NR

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NH	Number of Obs	1426719		1402085		1402085	
Z_NH	Mean	-0.1484	-0.0685	0.0047	0.0451	0.0056	0.0446
Z_NH	Variance	0.5889	0.9292	0.1803	0.2572	0.1868	0.2617
Z_NH	Number of Missing Obs	7138369		7250475		7250475	
Z_NH	99th Percentile	2.2000	2.1667	1.3256	1.1379	1.3432	1.1317
Z_NH	95th Percentile	0.5152	0.6260	0.3710	0.3082	0.3773	0.3189
Z_NH	90th Percentile	0.1650	0.2787	0.1414	0.1607	0.1441	0.1631
Z_NH	75th Percentile	0.0000	0.0586	0.0000	0.0585	0.0000	0.0568
Z_NH	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0017
Z_NH	25th Percentile	-0.4100	-0.1367	0.0000	-0.0197	-0.0057	-0.0208
Z_NH	10th Percentile	-1.0000	-1.0000	-0.1925	-0.0952	-0.1945	-0.1002
Z_NH	5th Percentile	-1.0000	-1.0000	-0.4398	-0.1869	-0.4413	-0.1937
Z_NH	1st Percentile	-1.0000	-1.0000	-1.0000	-0.7437	-1.0000	-0.7681
Z_NH	90th - 10th	1.1650	1.2787	0.3339	0.2560	0.3386	0.2633
Z_NH	75th - 25th	0.4100	0.1953	0.0000	0.0782	0.0057	0.0776
Z_NR	Number of Obs	624897		617327		617328	
Z_NR	Mean	-0.2288	-0.2032	-0.0225	0.0009	-0.0221	0.0008
Z_NR	Variance	0.3014	0.5088	0.0775	0.0803	0.0786	0.0803
Z_NR	Number of Missing Obs	7940191		8035233		8035232	
Z_NR	99th Percentile	1.0000	1.5641	0.8400	0.8299	0.8476	0.8421
Z_NR	95th Percentile	0.1982	0.3388	0.1905	0.2308	0.1949	0.2361
Z_NR	90th Percentile	0.0097	0.1205	0.0348	0.1034	0.0391	0.1047
Z_NR	75th Percentile	0.0000	0.0000	0.0000	0.0110	0.0000	0.0114
Z_NR	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Z_NR	25th Percentile	-0.7500	-0.8367	0.0000	-0.0185	0.0000	-0.0186
Z_NR	10th Percentile	-1.0000	-1.0000	-0.1350	-0.1094	-0.1368	-0.1109
Z_NR	5th Percentile	-1.0000	-1.0000	-0.3950	-0.2350	-0.3969	-0.2393
Z_NR	1st Percentile	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
Z_NR	90th - 10th	1.0097	1.1205	0.1698	0.2128	0.1759	0.2156
Z_NR	75th - 25th	0.7500	0.8367	0.0000	0.0295	0.0000	0.0300

Table F.35: Distribution of Bias Measures for County by 3-Digit SIC: Z_NS and Z_W2

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_NS	Number of Obs	1661635		1634628		1634628	
Z_NS	Mean	-0.1323	-0.0560	0.0057	0.0358	0.0063	0.0353
Z_NS	Variance	0.5781	0.9216	0.1556	0.1413	0.1588	0.1375
Z_NS	Number of Missing Obs	6903453		7017932		7017932	
Z_NS	99th Percentile	2.3250	2.3333	1.3300	1.0000	1.3510	1.0105
Z_NS	95th Percentile	0.5400	0.6760	0.3636	0.2881	0.3688	0.2926
Z_NS	90th Percentile	0.1700	0.2817	0.1368	0.1471	0.1390	0.1486
Z_NS	75th Percentile	0.0000	0.0550	0.0000	0.0574	0.0000	0.0559
Z_NS	50th Percentile	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013
Z_NS	25th Percentile	-0.3333	-0.1270	0.0000	-0.0171	-0.0015	-0.0182
Z_NS	10th Percentile	-1.0000	-1.0000	-0.1825	-0.0902	-0.1849	-0.0943
Z_NS	5th Percentile	-1.0000	-1.0000	-0.4200	-0.1835	-0.4237	-0.1883
Z_NS	1st Percentile	-1.0000	-1.0000	-1.0000	-0.6857	-1.0000	-0.7020
Z_NS	90th - 10th	1.1700	1.2817	0.3193	0.2373	0.3239	0.2429
Z_NS	75th - 25th	0.3333	0.1820	0.0000	0.0745	0.0015	0.0742
Z_W2	Number of Obs	4241913		4151260		4151260	
Z_W2	Mean	0.2299	-0.0862	0.1148	0.0105	0.1168	0.0103
Z_W2	Variance	128.7386	23.3782	18.4491	0.9740	18.6086	1.0140
Z_W2	Number of Missing Obs	4323175		4501300		4501300	
Z_W2	99th Percentile	5.7588	1.1238	1.9906	0.4802	2.0150	0.4887
Z_W2	95th Percentile	0.8824	0.2611	0.3270	0.1433	0.3338	0.1484
Z_W2	90th Percentile	0.2922	0.0987	0.1131	0.0698	0.1156	0.0726
Z_W2	75th Percentile	0.0000	0.0121	0.0000	0.0149	0.0000	0.0156
Z_W2	50th Percentile	0.0000	-0.0052	0.0000	0.0000	0.0000	0.0000
Z_W2	25th Percentile	-0.0831	-0.1463	0.0000	-0.0218	0.0000	-0.0233
Z_W2	10th Percentile	-1.0000	-0.7847	-0.0865	-0.0695	-0.0886	-0.0736
Z_W2	5th Percentile	-1.0000	-1.0000	-0.2286	-0.1213	-0.2319	-0.1277
Z_W2	1st Percentile	-1.0000	-1.0000	-1.0000	-0.3475	-1.0000	-0.3636
Z_W2	90th - 10th	1.2922	0.8834	0.1995	0.1392	0.2042	0.1462
Z_W2	75th - 25th	0.0831	0.1584	0.0000	0.0368	0.0000	0.0388

Table F.36: Distribution of Bias Measures for County by 3-Digit SIC: Z_W3 and Z_WFA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_W3	Number of Obs	3872482		3789339		3789339	
Z_W3	Mean	0.1750	-0.1000	0.1000	0.0065	0.1022	0.0062
Z_W3	Variance	131.3268	26.8136	22.7179	1.2574	23.4206	1.3034
Z_W3	Number of Missing Obs	4692606		4863221		4863221	
Z_W3	99th Percentile	4.7085	1.0192	1.6788	0.4503	1.7067	0.4585
Z_W3	95th Percentile	0.7693	0.2392	0.2985	0.1328	0.3057	0.1383
Z_W3	90th Percentile	0.2599	0.0880	0.1047	0.0642	0.1073	0.0670
Z_W3	75th Percentile	0.0000	0.0086	0.0000	0.0121	0.0000	0.0130
Z_W3	50th Percentile	0.0000	-0.0075	0.0000	0.0000	0.0000	0.0000
Z_W3	25th Percentile	-0.0852	-0.1657	0.0000	-0.0233	0.0000	-0.0247
Z_W3	10th Percentile	-1.0000	-0.8906	-0.0811	-0.0705	-0.0835	-0.0755
Z_W3	5th Percentile	-1.0000	-1.0000	-0.2139	-0.1294	-0.2171	-0.1354
Z_W3	1st Percentile	-1.0000	-1.0000	-1.0000	-0.3466	-1.0000	-0.3647
Z_W3	90th - 10th	1.2599	0.9786	0.1858	0.1347	0.1908	0.1424
Z_W3	75th - 25th	0.0852	0.1743	0.0000	0.0354	0.0000	0.0377
Z_WFA	Number of Obs	1438501		1410963		1410963	
Z_WFA	Mean	0.0333	-0.1507	0.1174	0.0069	0.1177	0.0070
Z_WFA	Variance	44.2546	8.9867	52.8114	8.9875	27.8320	4.9095
Z_WFA	Number of Missing Obs	7126587		7241597		7241597	
Z_WFA	99th Percentile	3.8646	1.3573	2.3173	0.8394	2.3529	0.8667
Z_WFA	95th Percentile	0.5963	0.2736	0.4000	0.2070	0.4093	0.2137
Z_WFA	90th Percentile	0.1681	0.0982	0.1315	0.0915	0.1349	0.0946
Z_WFA	75th Percentile	0.0000	0.0007	0.0000	0.0096	0.0000	0.0103
Z_WFA	50th Percentile	0.0000	-0.0257	0.0000	-0.0008	0.0000	-0.0017
Z_WFA	25th Percentile	-0.3378	-0.3557	0.0000	-0.0576	-0.0006	-0.0597
Z_WFA	10th Percentile	-1.0000	-1.0000	-0.1398	-0.1743	-0.1424	-0.1765
Z_WFA	5th Percentile	-1.0000	-1.0000	-0.3323	-0.2671	-0.3362	-0.2703
Z_WFA	1st Percentile	-1.0000	-1.0000	-1.0000	-0.8025	-1.0000	-0.8399
Z_WFA	90th - 10th	1.1681	1.0982	0.2713	0.2658	0.2773	0.2712
Z_WFA	75th - 25th	0.3378	0.3564	0.0000	0.0672	0.0006	0.0699

Table F.37: Distribution of Bias Measures for County by 3-Digit SIC: Z_WFS and Z_WH3

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_WFS	Number of Obs	1348132		1323883		1323883	
Z_WFS	Mean	0.5338	-0.0029	0.4361	0.1178	0.4401	0.1182
Z_WFS	Variance	1960.2673	248.4258	287.5078	56.3677	276.7465	53.4042
Z_WFS	Number of Missing Obs	7216956		7328677		7328677	
Z_WFS	99th Percentile	11.1642	2.8735	6.6563	1.7644	6.6935	1.7926
Z_WFS	95th Percentile	1.1881	0.5109	0.7729	0.3971	0.7802	0.4027
Z_WFS	90th Percentile	0.2830	0.1867	0.2305	0.1864	0.2318	0.1879
Z_WFS	75th Percentile	0.0000	0.0201	0.0000	0.0405	0.0000	0.0393
Z_WFS	50th Percentile	0.0000	-0.0078	0.0000	0.0000	0.0000	0.0000
Z_WFS	25th Percentile	-0.5036	-0.4313	0.0000	-0.0385	0.0000	-0.0398
Z_WFS	10th Percentile	-1.0000	-1.0000	-0.1681	-0.1560	-0.1695	-0.1584
Z_WFS	5th Percentile	-1.0000	-1.0000	-0.4487	-0.2909	-0.4518	-0.2961
Z_WFS	1st Percentile	-1.0000	-1.0000	-1.0000	-0.9395	-1.0000	-0.9610
Z_WFS	90th - 10th	1.2830	1.1867	0.3986	0.3424	0.4014	0.3463
Z_WFS	75th - 25th	0.5036	0.4515	0.0000	0.0790	0.0000	0.0791
Z_WH3	Number of Obs	1181426		1157733		1157733	
Z_WH3	Mean	-0.0255	-0.1722	0.0913	-0.0008	0.0940	-0.0003
Z_WH3	Variance	29.3017	6.3881	11.8115	2.4366	12.6458	2.6178
Z_WH3	Number of Missing Obs	7383662		7494827		7494827	
Z_WH3	99th Percentile	3.3219	1.2977	2.1048	0.8138	2.1455	0.8383
Z_WH3	95th Percentile	0.5639	0.2662	0.3933	0.2097	0.4030	0.2176
Z_WH3	90th Percentile	0.1616	0.0982	0.1336	0.0939	0.1375	0.0977
Z_WH3	75th Percentile	0.0000	0.0000	0.0000	0.0102	0.0000	0.0108
Z_WH3	50th Percentile	0.0000	-0.0312	0.0000	-0.0024	0.0000	-0.0034
Z_WH3	25th Percentile	-0.4239	-0.4016	0.0000	-0.0637	-0.0031	-0.0661
Z_WH3	10th Percentile	-1.0000	-1.0000	-0.1499	-0.1897	-0.1526	-0.1929
Z_WH3	5th Percentile	-1.0000	-1.0000	-0.3492	-0.2890	-0.3530	-0.2906
Z_WH3	1st Percentile	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
Z_WH3	90th - 10th	1.1616	1.0982	0.2836	0.2836	0.2901	0.2906
Z_WH3	75th - 25th	0.4239	0.4016	0.0000	0.0739	0.0031	0.0768

Table F.38: Distribution of Bias Measures for County by 3-Digit SIC: A and FA

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
A	Number of Obs	8358016	n/a	8443648	n/a	8443648	n/a
A	Mean	-0.0964	n/a	-0.0923	n/a	-0.0976	n/a
A	Variance	645.5399	n/a	55.9108	n/a	57.0365	n/a
A	Number of Missing Obs	207072	n/a	208912	n/a	208912	n/a
A	99th Percentile	10.0000	n/a	4.0000	n/a	4.0000	n/a
A	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
A	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
A	1st Percentile	-11.0000	n/a	-4.0000	n/a	-4.0000	n/a
A	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
A	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	Number of Obs	7943696	n/a	8025488	n/a	8025488	n/a
FA	Mean	0.0455	n/a	0.0563	n/a	0.0540	n/a
FA	Variance	106.8865	n/a	20.7861	n/a	21.2355	n/a
FA	Number of Missing Obs	621392	n/a	627072	n/a	627072	n/a
FA	99th Percentile	7.0000	n/a	3.0000	n/a	3.0000	n/a
FA	95th Percentile	1.0000	n/a	0.0000	n/a	0.0000	n/a
FA	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
FA	1st Percentile	-6.0000	n/a	-2.0000	n/a	-2.0000	n/a
FA	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FA	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.39: Distribution of Bias Measures for County by 3-Digit SIC: FJC and FJD

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJC	Number of Obs	7943696	n/a	8025488	n/a	4306800	n/a
FJC	Mean	0.0270	n/a	0.0339	n/a	0.0610	n/a
FJC	Variance	62.1825	n/a	19.3240	n/a	36.2677	n/a
FJC	Number of Missing Obs	621392	n/a	627072	n/a	4345760	n/a
FJC	99th Percentile	5.0000	n/a	2.0000	n/a	4.0000	n/a
FJC	95th Percentile	1.0000	n/a	0.0000	n/a	1.0000	n/a
FJC	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	5th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
FJC	1st Percentile	-4.0000	n/a	-2.0000	n/a	-3.0000	n/a
FJC	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	Number of Obs	7943696	n/a	8025488	n/a	4306800	n/a
FJD	Mean	0	n/a	0	n/a	0	n/a
FJD	Variance	25.4231	n/a	5.8924	n/a	11.2261	n/a
FJD	Number of Missing Obs	621392	n/a	627072	n/a	4345760	n/a
FJD	99th Percentile	3	n/a	1	n/a	2	n/a
FJD	95th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	5th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
FJD	1st Percentile	-4.0000	n/a	-2.0000	n/a	-4.0000	n/a
FJD	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.40: Distribution of Bias Measures for County by 3-Digit SIC: FJF and FS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FJF	Number of Obs	7943696	n/a	8025488	n/a	4306800	n/a
FJF	Mean	0.0854	n/a	0.0883	n/a	0.1662	n/a
FJF	Variance	60.8045	n/a	27.9424	n/a	52.7543	n/a
FJF	Number of Missing Obs	621392	n/a	627072	n/a	4345760	n/a
FJF	99th Percentile	6.0000	n/a	4.0000	n/a	7.0000	n/a
FJF	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
FJF	90th Percentile	0.0000	n/a	0.0000	n/a	1.0000	n/a
FJF	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FJF	5th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
FJF	1st Percentile	-5.0000	n/a	-2.0000	n/a	-4.0000	n/a
FJF	90th - 10th	0.0000	n/a	0.0000	n/a	1.0000	n/a
FJF	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	Number of Obs	7943696	n/a	8025488	n/a	8025488	n/a
FS	Mean	-0.0407	n/a	-0.0325	n/a	-0.0355	n/a
FS	Variance	54.3145	n/a	7.0139	n/a	7.1813	n/a
FS	Number of Missing Obs	621392	n/a	627072	n/a	627072	n/a
FS	99th Percentile	4.0000	n/a	2.0000	n/a	2.0000	n/a
FS	95th Percentile	1.0000	n/a	0.0000	n/a	0.0000	n/a
FS	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
FS	1st Percentile	-5.0000	n/a	-2.0000	n/a	-2.0000	n/a
FS	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
FS	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.41: Distribution of Bias Measures for County by 3-Digit SIC: FT and H

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
FT	Number of Obs	7738912	n/a	7690848	n/a	7690848	n/a
FT	Mean	0.0179	n/a	0.0200	n/a	0.0203	n/a
FT	Variance	1.5967	n/a	1.3107	n/a	1.2922	n/a
FT	Number of Missing Obs	826176	n/a	961712	n/a	961712	n/a
FT	99th Percentile	0.5300	n/a	0.5000	n/a	0.5000	n/a
FT	95th Percentile	0.2800	n/a	0.1300	n/a	0.1260	n/a
FT	90th Percentile	0.1100	n/a	0.0300	n/a	0.0300	n/a
FT	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
FT	10th Percentile	-0.0700	n/a	0.0000	n/a	-0.0010	n/a
FT	5th Percentile	-0.2500	n/a	-0.0500	n/a	-0.0480	n/a
FT	1st Percentile	-0.5000	n/a	-0.4600	n/a	-0.4530	n/a
FT	90th - 10th	0.1800	n/a	0.0300	n/a	0.0310	n/a
FT	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	Number of Obs	7737088	n/a	7817472	n/a	7817472	n/a
H	Mean	-0.0963	n/a	-0.0921	n/a	-0.0971	n/a
H	Variance	461.8109	n/a	48.9267	n/a	50.1972	n/a
H	Number of Missing Obs	828000	n/a	835088	n/a	835088	n/a
H	99th Percentile	9.0000	n/a	3.0000	n/a	3.0000	n/a
H	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
H	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
H	1st Percentile	-9.0000	n/a	-4.0000	n/a	-4.0000	n/a
H	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
H	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.42: Distribution of Bias Measures for County by 3-Digit SIC: H3 and JC

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
H3	Number of Obs	7322384	n/a	7398864	n/a	7398864	n/a
H3	Mean	0.0448	n/a	0.0561	n/a	0.0540	n/a
H3	Variance	79.4919	n/a	18.8761	n/a	19.3719	n/a
H3	Number of Missing Obs	1242704	n/a	1253696	n/a	1253696	n/a
H3	99th Percentile	6.0000	n/a	3.0000	n/a	3.0000	n/a
H3	95th Percentile	1.0000	n/a	0.0000	n/a	0.0000	n/a
H3	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
H3	1st Percentile	-5.0000	n/a	-2.0000	n/a	-2.0000	n/a
H3	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
H3	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	Number of Obs	8151648	n/a	8235232	n/a	4731572	n/a
JC	Mean	-0.0541	n/a	-0.0505	n/a	-0.0930	n/a
JC	Variance	45.2700	n/a	14.1231	n/a	25.0402	n/a
JC	Number of Missing Obs	413440	n/a	417328	n/a	3920988	n/a
JC	99th Percentile	4.0000	n/a	2.0000	n/a	3.0000	n/a
JC	95th Percentile	1.0000	n/a	0.0000	n/a	1.0000	n/a
JC	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	5th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
JC	1st Percentile	-5.0000	n/a	-3.0000	n/a	-4.0000	n/a
JC	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JC	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.43: Distribution of Bias Measures for County by 3-Digit SIC: JD and JF

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
JD	Number of Obs	8151648	n/a	8235232	n/a	4731572	n/a
JD	Mean	-0.0454	n/a	-0.0406	n/a	-0.0745	n/a
JD	Variance	48.8682	n/a	8.4236	n/a	14.6738	n/a
JD	Number of Missing Obs	413440	n/a	417328	n/a	3920988	n/a
JD	99th Percentile	4.0000	n/a	2.0000	n/a	3.0000	n/a
JD	95th Percentile	1.0000	n/a	0.0000	n/a	1.0000	n/a
JD	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	5th Percentile	-1.0000	n/a	0.0000	n/a	-1.0000	n/a
JD	1st Percentile	-5.0000	n/a	-2.0000	n/a	-4.0000	n/a
JD	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JD	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	Number of Obs	8151648	n/a	8235232	n/a	4731572	n/a
JF	Mean	-0.0086	n/a	-0.0097	n/a	-0.0182	n/a
JF	Variance	54.5757	n/a	21.0267	n/a	37.0468	n/a
JF	Number of Missing Obs	413440	n/a	417328	n/a	3920988	n/a
JF	99th Percentile	6.0000	n/a	3.0000	n/a	5.0000	n/a
JF	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
JF	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	5th Percentile	-1.0000	n/a	-1.0000	n/a	-1.0000	n/a
JF	1st Percentile	-6.0000	n/a	-3.0000	n/a	-6.0000	n/a
JF	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
JF	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.44: Distribution of Bias Measures for County by 3-Digit SIC: R and S

Relevant Variable	Statistic	Raw No Establishment Weight Applied Unweighted Bias	Weighted Bias	Raw Establishment Weight Applied Unweighted Bias	Weighted Bias	Fuzzed Establishment Weight Applied Unweighted Bias	Weighted Bias
R	Number of Obs	7737088	n/a	7817472	n/a	7817472	n/a
R	Mean	-0.0086	n/a	-0.0049	n/a	-0.0068	n/a
R	Variance	23.6475	n/a	3.1220	n/a	3.0977	n/a
R	Number of Missing Obs	828000	n/a	835088	n/a	835088	n/a
R	99th Percentile	2.0000	n/a	1.0000	n/a	1.0000	n/a
R	95th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	5th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	1st Percentile	-2.0000	n/a	-1.0000	n/a	-1.0000	n/a
R	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
R	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	Number of Obs	8358720	n/a	8444144	n/a	8444144	n/a
S	Mean	-0.0872	n/a	-0.0820	n/a	-0.0864	n/a
S	Variance	664.7815	n/a	43.1072	n/a	43.9474	n/a
S	Number of Missing Obs	206368	n/a	208416	n/a	208416	n/a
S	99th Percentile	10.0000	n/a	4.0000	n/a	4.0000	n/a
S	95th Percentile	1.0000	n/a	1.0000	n/a	1.0000	n/a
S	90th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	75th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	25th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	10th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	5th Percentile	-1.0000	n/a	0.0000	n/a	0.0000	n/a
S	1st Percentile	-11.0000	n/a	-4.0000	n/a	-4.0000	n/a
S	90th - 10th	0.0000	n/a	0.0000	n/a	0.0000	n/a
S	75th - 25th	0.0000	n/a	0.0000	n/a	0.0000	n/a

Table F.45: Distribution of Bias Measures for County by 3-Digit SIC: Z_dWA and Z_dWS

Relevant Variable	Statistic	Raw No Establishment Weight Applied		Raw Establishment Weight Applied		Fuzzed Establishment Weight Applied	
		Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias	Unweighted Bias	Weighted Bias
Z_dWA	Number of Obs	3057395	n/a	2695426	n/a	2695426	n/a
Z_dWA	Mean	106.3887	n/a	127.3084	n/a	130.4103	n/a
Z_dWA	Variance	329494866.6676	n/a	353584649.1531	n/a	441335175.7838	n/a
Z_dWA	Number of Missing Obs	5507693	n/a	5957134	n/a	5957134	n/a
Z_dWA	99th Percentile	3975.0000	n/a	2998.0000	n/a	3019.0000	n/a
Z_dWA	95th Percentile	1329.0000	n/a	893.0000	n/a	894.0000	n/a
Z_dWA	90th Percentile	716.0000	n/a	427.0000	n/a	425.0000	n/a
Z_dWA	75th Percentile	171.0000	n/a	49.0000	n/a	49.0000	n/a
Z_dWA	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWA	25th Percentile	-38.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWA	10th Percentile	-480.0000	n/a	-129.0000	n/a	-128.0000	n/a
Z_dWA	5th Percentile	-1000.0000	n/a	-376.0000	n/a	-376.0000	n/a
Z_dWA	1st Percentile	-3211.0000	n/a	-1650.0000	n/a	-1655.0000	n/a
Z_dWA	90th - 10th	1196.0000	n/a	556.0000	n/a	553.0000	n/a
Z_dWA	75th - 25th	209.0000	n/a	49.0000	n/a	49.0000	n/a
Z_dWS	Number of Obs	3064986	n/a	2700666	n/a	2700666	n/a
Z_dWS	Mean	-106.9739	n/a	-130.6706	n/a	-133.9807	n/a
Z_dWS	Variance	145317324.7758	n/a	161409167.9228	n/a	194012846.6077	n/a
Z_dWS	Number of Missing Obs	5500102	n/a	5951894	n/a	5951894	n/a
Z_dWS	99th Percentile	3333.0000	n/a	1676.0000	n/a	1681.0000	n/a
Z_dWS	95th Percentile	1051.0000	n/a	416.0000	n/a	414.0000	n/a
Z_dWS	90th Percentile	506.0000	n/a	154.0000	n/a	154.0000	n/a
Z_dWS	75th Percentile	48.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWS	50th Percentile	0.0000	n/a	0.0000	n/a	0.0000	n/a
Z_dWS	25th Percentile	-150.0000	n/a	-36.0000	n/a	-37.0000	n/a
Z_dWS	10th Percentile	-737.0000	n/a	-405.0000	n/a	-403.0000	n/a
Z_dWS	5th Percentile	-1403.0000	n/a	-910.0000	n/a	-909.0000	n/a
Z_dWS	1st Percentile	-4224.0000	n/a	-3208.0000	n/a	-3239.0000	n/a
Z_dWS	90th - 10th	1243.0000	n/a	559.0000	n/a	557.0000	n/a
Z_dWS	75th - 25th	198.0000	n/a	36.0000	n/a	37.0000	n/a

Bibliography

Abowd, J., F. Kramarz, and D. Margolis (1999, March). High wage workers and high wage firms. *Econometrica*.

Abowd, J. M., J. C. Haltiwanger, and J. I. Lane (2004). Integrated longitudinal employee-employer data for the United States. *American Economic Review 94*(2).

Abowd, J. M., F. Kramarz, and S. Roux (2006, June). Wages, mobility and firm performance: Advantages and insights from using matched worker-firm data. *The Economic Journal 116*, 245–285.

Abowd, J. M., B. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock (2005). The LEHD infrastructure files and the creation of the quarterly workforce indicators. LEHD Technical Paper TP-2006-01.

Baker, G., M. Gibbs, and B. Holmstrom (1995). The wage policy of a firm. *The Quarterly Journal of Economics 107*, 921–957.

Bronars, S. G. and M. Famulari (1997). Wage, tenure, and wage growth variation within and across establishments. *Journal of Labor Economics 15*(2), 285–317.

Brown, C. and J. Medoff (1989, October). The employer size-wage effect. *The Journal of Political Economy 97*(5), 1027–1059.

Cappelli, P. and D. Neumark (2001). Do "High Performance" work practices improve establishment-level outcomes? *Industrial Labor Relations Review 54*(4), 737–775.

Cardoso, A. (1999). Firms' wage policies and the rise in labor market inequality: The case of Portugal. *Industrial and Labor Relations Review 53*(1), 87–102.

Cardoso, A. (2000, July). Wage differentials across firms: An application of multilevel modelling. *Journal of Applied Econometrics*, 343–354.

Cox, L. H. and L. V. Zayatz (1993). An agenda for research in statistical disclosure limitation. Technical report, U.S. Bureau of the Census. Statistical Report Series LVZ93/01.

Davis, S. J. and J. Haltiwanger (1995). Employer size and the wage structure in U.S. manufacturing. NBER Working Paper 5393.

Evans, T., L. Zayatz, and J. Slanta (1998). Using noise for disclosure limitation of establishment tabular data. *Journal of Official Statistics*.

Gilmore, A. R., R. Thompson, and B. R. Cullis (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics 51*(4), 1440–1450.

Goldstein, H. (1995). *Multilevel Statistical Models*. Kendall's Library of Statistics. Halsted Press.

Groshen, E. (1991). Sources of intra-industry wage dispersion: How much do employers matter? *Quarterly Journal of Economics 106*(3), 869–884.

Groshen, E. and D. Levine (1998, July). The rise and decline of U.S. internal labor markets. Federal Reserve Bank of New York Research Paper 9819.

Ichniowski, C. and K. Shaw (2003). Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives 17*(1), 155–180.

Lazear, E. (2000). The future of personnel economics. *The Economic Journal 110*, 611–639.

Lazear, E. (2003). Firm-specific human capital: A skill weights approach. NBER Working Paper 9679.

Little, R. J. A. and D. B. Rubin (1986). *Statistical Analysis with Missing Data*. John Wiley and Sons, Inc.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, Linear, and Mixed Models*. John Wiley and Sons, Inc.

Prendergast, C. (1999, March). The provision of incentives in firms. *Journal of Economic Literature*, 7–63.

Raudenbush, S. and A. Bryk (1986, January). A hierarchical model for studying school effects. *Sociology of Education*, 1–17.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse Surveys*. John Wiley and Sons, Inc.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. John Wiley and Sons, Inc.

Woodcock, S. D. (2003, July). Heterogeneity and learning in labor markets. Working paper.