11. Inverse probability weighting is a flexible method of covariate adjustment developed primarily to draw causal inferences about the effects of treatments in observational studies (Robins, Hernan, and Brumback 2000). This method has other applications, however. At a basic level, it is simply a form of direct standardization—a widely-used technique in demography—that treats the marginal distribution of covariates as the standard to which classes (or treatment groups) are transformed via weighting. This form of standardization addresses the following type of counterfactual question: if, for example, workers and proprietors had the same educational distribution as the overall population at each time point, how would these groups differ in terms of mean income over time?

12. The IP-weighted local linear regression estimator is given by $\hat{g}_j(t) = \sum_{i=1}^{n_j} a_{ij}^w y_{ij} = \sum_{i=1}^{n_j}\left(\frac{w_{ij}}{\sum_{i=1}^{n_j} w_{ij}}\right)\frac{\left(S_2^w(t)-S_1^w(t)(t_{ij}-t)\right)K_b(t_{ij}-t)}{S_2^w(t)S_0^w(t)-S_1^w(t)^2} y_{ij}$, where $S_v^w(t)$ is equal to $\sum_{i=1}^{n_j}\left(\frac{w_{ij}}{\sum_{i=1}^{n_j} w_{ij}}\right)(t_{ij}-t)^v K_b(t_{ij}-t)$ and $K_b(t_{ij}-t)$ is defined as previously.

13. The occupational proxy measure of class is subject to known error. For example, it classifies all employed doctors as managers, even though only about 60 percent of doctors report that their job involves supervisory responsibilities over subordinate workers (see Part B of the Online Supplement). If more women and minorities have become doctors over time, but this increased occupational representation is concentrated in positions that lack institutionalized authority in the workplace, measurement error would become more strongly correlated with these status groups over time, and the CPS would conflate increasing occupational integration with increasing integration across relational class boundaries, even if the later had not actually occurred (i.e., if white male doctors had continued to dominate positions of authority within that occupation, despite the influx of women and minorities).

REFERENCES

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in US Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90:300-23.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118:1279-333.

Baker, Dean and Lawrence Mishel. 1995. "Profits Up, Wages Down: Worker Losses Yield Big Gains for Business." Washington, DC: Economic Policy Institute.

Bakunin, Mikhail. 1953a. "Power and Authority." Pp. 48-55 in *The Political Philosophy of Bakunin*, edited by G. P. Maximoff. New York: The Free Press.

Bakunin, Mikhail. 1953b. "The Present Economic Regime." Pp. 182-188 in *The Political Philosophy of Bakunin*, edited by G. P. Maximoff. New York: The Free Press.

Bakunin, Mikhail. 1972. "The International and Karl Marx." Pp. 286-320 in *Bakunin On Anarchy: Selected Works by the Activist-founder of World Anarchism*, edited by S. Dolgoff. New York: Alfred A. Knopf.

Bakunin, Mikhail. 1992. *The Basic Bakunin*. New York: Prometheus Books.

Bakunin, Mikhail. 2013. *Marxism, Freedom, and the State*. Whitefish, MT: Kessinger.

Bernhardt, Annette, Martina Morris, and Mark S. Handcock. 1995. "Womens Gains or Mens Losses: A Closer Look at the Shrinking Gender-Gap in Earnings." *American Journal of Sociology* 101:302-28.

Bluestone, Barry and Bennett Harrison. 1982. *The Deindustrialization of America: Plant Closings, Community Abandonment, and the Dismantling of Basic Industry*. New York: Basic Books.

Breen, Richard. 2005. "Foundations of a Neo-Weberian Class Analysis." Pp. 31-50 in *Approaches to Class Analysis*, edited by E. O. Wright. Cambridge: Cambridge University Press.

Cancio, A. Silvia, T. David Evans, and Jr. Maume, David J. 1996. "Reconsidering the Declining Significance of Race: Racial Differences in Early Career Wages." *American Sociological Review* 61:541-56.

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20:733-83.

Clark, Terry N. and Seymour M. Lipset. 1991. "Are Social Classes Dying?" *International Sociology* 6:397-410.

Dahrendorf, Ralf. 1959. *Class and Class Conflict in Industrial Society*. Stanford, CA: Stanford University Press.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64:1001-44.

DiNardo, John and Justin L. Tobias. 2001. "Nonparametric Density and Regression Estimation." *Journal of Economic Perspectives* 15:11-28.

Dumenil, Gerard and Dominique Levy. 2004a. *Capital Resurgent: Roots of the Neoliberal Revolution*. Cambridge, MA: Harvard University Press.

Dumenil, Gerard and Dominique Levy. 2004b. "Neoliberal Income Trends: Wealth, Class and Ownership in the USA." *New Left Review*:105-133.

Durkheim, E. 1984. *The Division of Labor in Society*. New York: The Free Press.

Efron, Bradley and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Oxford University Press.

Foster, John Bellamy, Robert W. McChesney, and R. Jamil Jonna. 2011. "Monopoly and Competition in 21st Century Capitalism." *Monthly Review* 62:1-30.

Frydman, Carola and Dirk Jenter. 2010. "CEO Compensation." *Annual Review of Financial Economics* 2:75-102.

Grusky, David B. 2005. "Foundations of a Neo-Durkheimian Class Analysis." Pp. 51-81 in *Approaches to Class Analysis*, edited by E. O. Wright. Cambridge: Cambridge University Press.

Grusky, David B. and Jesper B. Sorensen. 1998. "Can Class Analysis Be Salvaged?" *American Journal of Sociology* 103:1187-234.

Halaby, Charles N. and David L. Weakliem. 1993. "Ownership and Authority in the Earnings Function: Nonnested Tests of Alternative Specifications." *American Sociological Review* 58:16-30.

Hao, Lingxin and Daniel Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks, CA: Sage.

Harrison, Bennett and Barry Bluestone. 1988. *The Great U-Turn: Corporate Restructuring and the Polarizing of America*. New York: Basic Books.

Harvey, David. 2005. *A Brief History of Neoliberalism*. Oxford, UK: Oxford University Press.

Hout, Michael. 2004. "GSS Methodological Report 101: Getting the Most Out of the GSS Income Measures." National Opinion Research Center, Chicago.

Hout, Michael, Clem Brooks, and Jeff Manza. 1993. "The Persistence of Classes in Postindustrial Societies." *International Sociology* 8:259-77.

Juhn, Chinhui. 1999. "Wage Inequality and Demand for Skill: Evidence from Five Decades." *Industrial & Labor Relations Review* 52:424-43.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101:410-42.

Kalleberg, Arne L. and Larry J. Griffin. 1980. "Class, Occupation, and Inequality in Job Rewards." *American Journal of Sociology* 85:731-68.

Kerbo, Harold R. 2009. *Social Stratification and Inequality: Class Conflict in Historical, Comparative, and Global Perspective*. New York: McGraw-Hill.

Kim, ChangHwan and Arthur Sakamoto. 2008. "The Rise of Intra-Occupational Wage Inequality in the United States, 1983 to 2002." *American Sociological Review* 73:129-57.

King, Miriam, Steven Ruggles, J. Trent Alexander, Sarah Flood, Katie Genadek, Matthew B. Schroeder, Brandon Trampe, and Rebecca Vick. 2010. "Integrated Public Use Microdata Series, Current Population Survey: Version 3.0 [machine-readable database]." Minneapolis, MN: Minnesota Population Center.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15:143-56.

Kristal, Tali. 2013. "The Capitalist Machine: Computerization, Workers' Power, and the Decline in Labor's Share within U.S. Industries." *American Sociological Review* 78:361-89.

Kropotkin, Peter. 1970. *Kropotkin's Revolutionary Pamphlets*. New York: Dover.

Kvam, Paul H. and B. Vidakovic. 2007. *Nonparametric Statistics with Applications to Science and Engineering*. Hoboken, NJ: Wiley & Sons.

Larrimore, Jeff, Richard V. Burkhauser, Shuaizhang Feng, and Laura Zayatz. 2008. "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976-2007), Working Paper 13941." Cambridge, MA: National Bureau of Economic Research.

Lynn, Barry C. 2011. *Cornered: The New Monopoly Capitalism and the Economics of Destruction*. Hoboken, NJ: John Wiley & Sons.

Marx, Karl. 1971. *Capital*, *Vol. 1*. London: George Allen and Unwin.

Marx, Karl. 1978a. "Capital, Vol. 3." Pp. 439-42 in *The Marx-Engels Reader*, edited by R. C. Tucker. New York: W.W. Norton & Company.

Marx, Karl. 1978b. "Manifesto of the Communist Party." Pp. 469-500 in *The Marx-Engels Reader*, edited by R. C. Tucker. New York: W. W. Norton and Company.

McCall, Leslie and Christine Percheski. 2010. "Income Inequality: New Trends and Research Directions." *Annual Review of Sociology* 36:329-47.

McKay, Iain. 2011. "Introduction." Pp. 1-51 in *Property is Theft: A Pierre-Joseph Proudhon Anthology*, edited by I. McKay. Oakland, CA: AK Press.

Mizruchi, Mark S. 2013. *The Fracturing of the American Corporate Elite*. Cambridge: Harvard University Press.

Morgan, Stephen L. and Youngjoo Cha. 2007. "Rent and the Evolution of Inequality in Late Industrial United States." *American Behavioral Scientist* 50:677-701.

Morgan, Stephen L. and Zun Tang. 2007. "Social Class and Workers' Rent, 1983-2001." *Research in Social Stratification and Mobility* 25:273-93.

Morris, Martina and Bruce Western. 1999. "Inequality in Earnings at the Close of the Twentieth Century." *Annual Review of Sociology* 25:623-57.

Mouw, Ted and Arne L. Kalleberg. 2010. "Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s." *American Sociological Review* 75:402-31.

Nisbet, Robert. 1959. "The Decline and Fall of Social Class." *Pacific Sociological Review* 2:11-28.

Pakulski, J. and M. Waters. 1996. "The Death of Class." Pp. 1022-1029 in *Social Stratification: Class, Race, and Gender in Sociological Perspective*, edited by D. B. Grusky. Boulder, CO: Westview.

Pakulski, Jan. 2005. "Foundations of Post-Class Analysis." Pp. 152-79 in *Approaches to Class Analysis*, edited by E. O. Wright. Cambridge: Cambridge University Press.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118:1-39.

Piketty, Thomas and Emmanuel Saez. 2006. "The Evolution of Top Incomes: A Historical and International Perspective." *American Economic Review* 96:200-5.

Proudhon, Pierre-Joseph. 1994. *What is Property?* Cambridge: Cambridge University Press.

Proudhon, Pierre-Joseph. 2011a. *Property Is Theft: A Pierre-Joseph Proudhon Reader* Oakland, CA: AK Press.

Proudhon, Pierre-Joseph. 2011b. "System of Economic Contradictions." Pp. 167-254 in *Property is Theft: A Pierre-Joseph Proudhon Anthology*, edited by I. McKay. Oakland, CA: AK Press.

Robins, J. M., M. A. Hernan, and B. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11:550-560.

Robinson, Robert V. and Jonathan Kelley. 1979. "Class as Conceived by Marx and Dahrendorf: Effects on Income Inequality and Politics in the United States and Great Britain." *American Sociological Review* 44:38-58.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.

Schmidt, Michael and Lucien van der Walt. 2009. *Black Flame: The Revolutionary Class Politics of Anarchism and Syndicalism* Oakland, CA: AK Press.

Schmitt, John and Ben Zipperer. 2009. "Dropping the Axe: Trends in Illegal Firings During Union Election Campaigns, 1951-2007." Center for Economic and Policy Research, Washington, DC.

Smith, Tom W., Peter V. Marsden, Michael Hout, and Jibum Kim. 2011. "General Social Surveys, 1972-2010 [machine-readable data file]." Chicago: National Opinion Research Center.

Thomas, Melvin E. 1993. "Race, Class, and Personal Income: An Empirical Test of the Declining Significance of Race." *Social Problems* 40:328-42.

Thorndike, R. L. 1942. "Two Screening Tests of Verbal Intelligence." *Journal of Applied Psychology* 26:128-135.

Useem, Michael. 1984. *The Inner Circle: Large Corporations and the Rise of Business Political Activity in the U.S. and U.K.* New York: Oxford University Press.

Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Weeden, Kim A. and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111:141-212.

Weeden, Kim A., Young-Mi Kim, Matthew Di Carlo, and David B. Grusky. 2007. "Social Class and Earnings Inequality." *American Behavioral Scientist* 50:702-36.

Western, Bruce. 2009. "Variance Function Regressions for Studying Inequality." *Sociological Methodology* 39:293-326.

Wright, Erik O. 1978. "Race, Class, and Income Inequality." *American Journal of Sociology* 83:1368-97.

Wright, Erik O. 1979. *Class Structure and Income Determination*. New York: Academic.

Wright, Erik O. 1984. "A General Framework for the Analysis of Class Structure." *Politics & Society* 13:383-423.

Wright, Erik O. 1985. *Classes*. London: Verso.

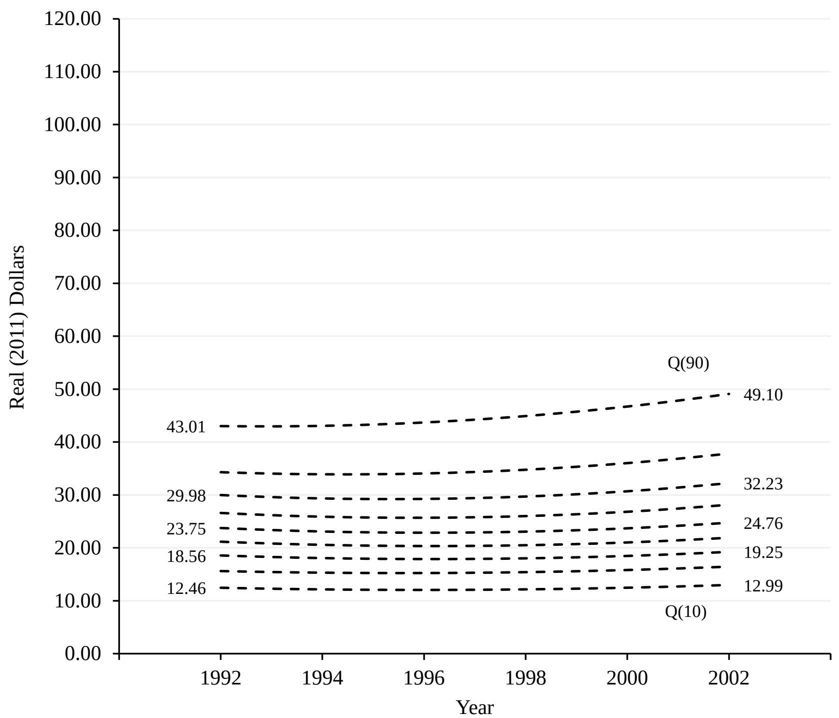
Wright, Erik O. 1994. *Interrogating Inequality: Essays on Class Analysis, Socialism and Marxism*. London: Verso.

Wright, Erik O. 1996. "The Continuing Relevance of Class Analysis." *Theory and Society* 25:693-716.

Wright, Erik O. 1997. *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Wright, Erik O. and Luca Perrone. 1977. "Marxist Class Categories and Income Inequality." *American Sociological Review* 42:32-55.

Yatchew, Adonis. 1998. "Nonparametric Regression Techniques in Economics." *Journal of Economic Literature* 36:669-721.



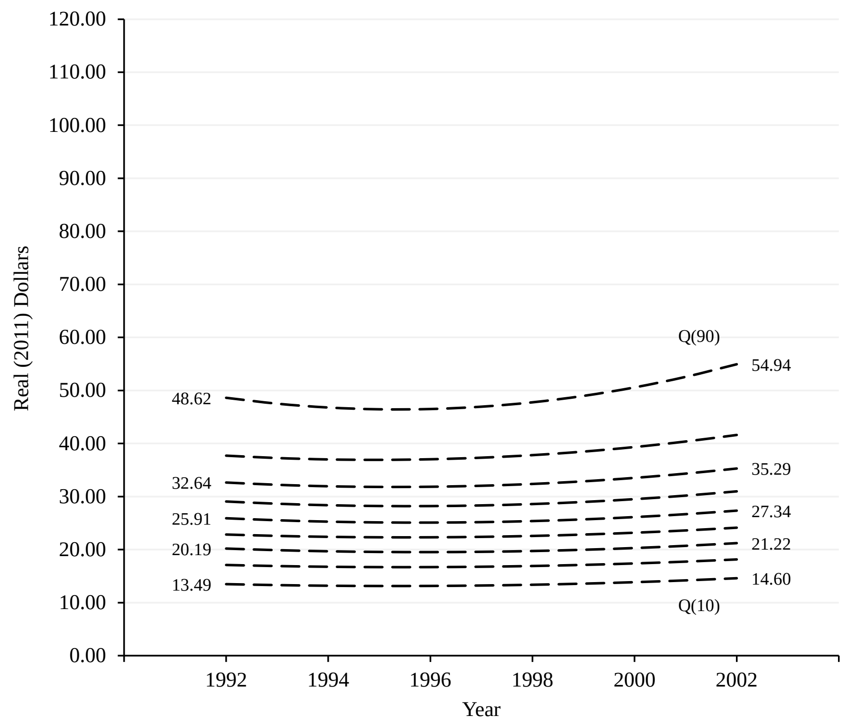
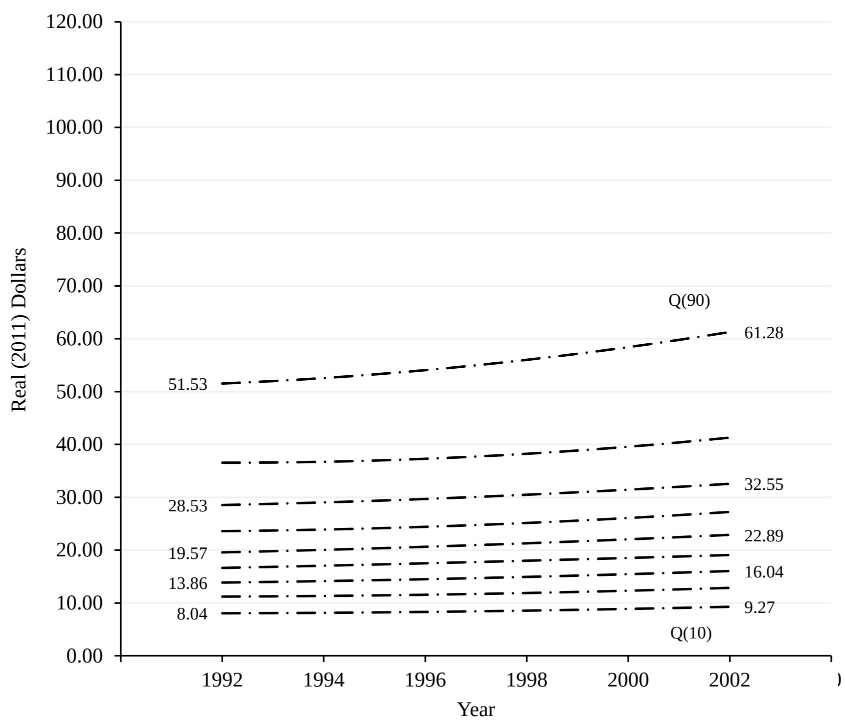
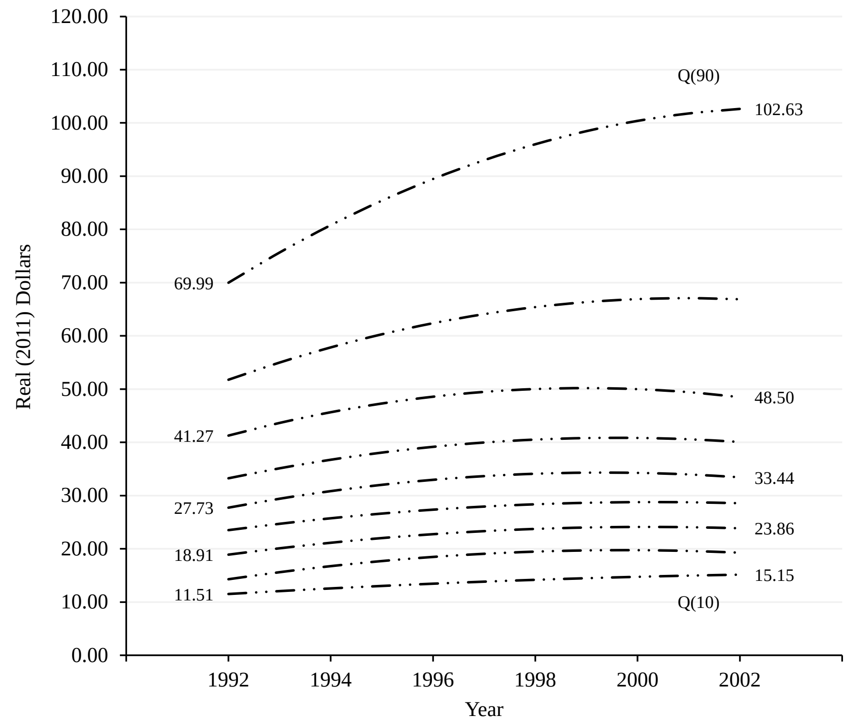


Figure A7 continued

G. Small Proprietors



H. Large Proprietors



Notes: Sample includes respondents who are 18 to 65 years old and work either full- or part-time in the 1992 to 2002 CPS waves. Results are combined estimates from 5 multiple imputation datasets and are based on quantile regression models with quadratic functions of time.

Part B: Measuring Class with Occupational Data from the CPS

This section describes and evaluates the procedures used to assign individuals to classes using occupational categories and self-employment data from the CPS. With this approach, census occupational categories are used as a proxy measure for supervisory and managerial responsibilities at work. Respondents in occupations that typically involve supervisory or managerial responsibilities are classified as proprietors if they are self-employed and as managers if they are employed by someone else. Conversely, respondents in occupations that do not typically involve supervisory or managerial responsibilities are classified as independent producers if they are self-employed and workers if they are employed by someone else.

To determine whether occupations typically involve supervisory or managerial responsibilities, highly specific occupational categories are first aggregated into larger groups to yield sufficiently large sample sizes. Then, these occupational groups are cross-tabulated with respondent reports of supervisory responsibilities from the GSS, which collects data on occupation and directly asks about various types of job responsibilities. This cross-tabulation is presented in the left-hand columns of Table B1, with occupational groups sorted in descending order based on the percentage of respondents who report that they supervise others at work.

Occupational groups in which more than 60 percent of respondents reported that they supervise others at work are defined as typically involving supervisory or managerial responsibilities. Respondents in these occupations are classified as either proprietors or managers, depending on their self-employment status. Occupational groups in which less than 50 percent of respondents reported supervising others at work are defined as not typically involving supervisory or managerial responsibilities. Respondents in these occupations are classified as either independent producers or workers, depending on their self-employment status. Note that there are not any occupational groups for which the percentage of respondents with supervisory responsibilities is entirely ambiguous (i.e., between 50 and 60 percent). The exact coding of the occupational proxy measure used in the main text is documented in the middle and right-hand columns of Table B1, which provide a breakdown of detailed occupational codes next to their class assignments.

This measurement strategy is subject to known error. It misclassifies respondents who have control over the work activities of others but are in occupations that do not *typically* involve supervisory or managerial responsibilities (e.g., a self-employed "carpenter" that owns a large construction company is classified as an independent producer rather than a proprietor). It also misclassifies respondents who do not have any control over the work activities of others but are in an occupation that does *typically* involve supervisory or managerial responsibilities (e.g., an employed “real estate agent” without a subordinate staff or any decision latitude at work is classified as a manager rather than a worker). Despite these known limitations, occupational categories still capture important information about supervisory and managerial responsibilities, and thus in most cases, these data can be used together with self-employment information to accurately assign individuals to classes.

For a certain subset of occupations, however, this measurement strategy is particularly prone to error because these occupational categories do not closely reflect the social relations of production. Specifically, for engineering, skilled trade, and farming occupations, Table B1 shows that between 40 and 50 percent of respondents have supervisory or managerial responsibilities at work. This indicates that many individuals in these occupations have an inaccurate class assignment from the proxy measurement procedure. These results suggest that self-employed carpenters, for example, likely include both owners of large home building companies as well as small independent contractors. Similarly, carpenters that report working for someone else are likely composed of many regular employees with no control over the work activities of others as well as a large number of project managers directing construction at particular worksites. Engineers, skilled tradespersons, and farmers enter the social relations of production in a variety of different ways, making it particularly difficult to assign them to a relational class location with a high level of accuracy. Because of the inaccuracy associated with measuring the relational class position of engineers, skilled tradespersons, and farmers, this section evaluates properties of the primary occupational proxy measure on which results presented in the main text are based as well as an alternate proxy measure that additionally classifies respondents in these occupations as proprietors or managers (rather than independent producers or workers), depending on their self-employment status.

Table B2 contains static estimates of class size, separately by the different measures of described here. These estimates come from the 1994 to 2010 waves of GSS, which is the period for which all three class measures can be obtained in this survey. Results indicate that both the primary and alternate occupational proxy measures of class tend to overstate the number of workers and understate the number of managers. Furthermore, estimates of the relative size of different strata within classes indicate that this pattern of underestimation is primarily due to the occupational proxy measures misclassifying a nontrivial number of low-level managers as workers. These misclassified low-level managers are likely respondents in non-managerial occupations that nevertheless have supervisory responsibilities at the point of production.

Table B2 also provides misclassification rates, which give the percentage of respondents classified differently under the direct versus occupational proxy measures of class. These rates indicate that the primary occupational proxy measure assigns about 34 percent of respondents to an incorrect class position and about 40 percent of respondents to an incorrect class stratum. The alternate occupational proxy measure has slightly higher misclassification rates, as expected.

Table B3 presents results from static models of income fit to GSS data from 1994 to 2010, separately by the different class measures. These results help to assess whether class income differences based on the occupational proxy measures are consistent with those based on the direct measure of class. Table B3 indicates that estimates of class income differences based on the primary occupational proxy measure are reasonably consistent with those based on the direct measure. The only notable difference is that, compared to the direct measure, the primary occupational proxy measure overstates mean income for low-level managers. Taken together with the disparate estimates of class size documented in Table B2, this suggests that the main limitation of the primary occupational proxy measure is that it misclassifies a number of individuals with low-level supervisory responsibilities in nonmanagerial production and service occupations. As a result, it understates the size of the managerial class, and by disproportionately capturing respondents with more extensive authority at work, it also overstates mean income for the managerial class. Thus, when interpreting estimates from the CPS, it is important to keep these measurement limitations in mind.

Table B3 additionally indicates that the alternate occupational proxy measure considered in this section not only provides somewhat inaccurate income estimates for managers but also underestimates mean income for proprietors. Based on these results, the primary occupational proxy measure that classifies engineers, skilled tradespersons, and farmers as independent producers or workers is preferred over the alternate measure that classifies respondents in these occupations as proprietors or managers, depending on their self-employment status.

The highest level of confidence in results from the CPS can be achieved if it is also possible to demonstrate that the primary substantive conclusions of this study are insensitive to the choice of proxy measure. Figures B1 to B5 present results from parallel analyses of the CPS based on the primary and the alternate occupational proxy measures of class. Consistent with static income models fit to data from the GSS, these figures show that the alternate occupational proxy measure understates mean income for proprietors and especially for small proprietors. Despite these differences, the two proxy measures provide similar estimates of temporal change in mean income and ultimately justify similar substantive conclusions about trends in class income inequality since the early 1980s. Both the primary and alternate occupational proxy measures indicate that total class inequality has increased by at least 50 percent between 1983 and the mid-2000s. In sum, the supplementary analyses presented here indicate that occupational data provide a reasonable proxy measure for the relational definition class outlined in this study; that the primary occupational proxy measure that classifies engineers, skilled tradespersons, and farmers as independent producers or workers performs better than an alternate proxy measure that classifies these occupations as proprietors or managers; and that notwithstanding differences in performance between the primary and alternate proxy measure, both measures justify similar substantive conclusions about broad trends in class income differences.

Table B.1. Occupational Proxy Measure of Class

Occupational group	Pct w/ Auth in GSS	Proxy Coding		1980/1990 COCs	2000 COCs
		Self-employed	Employed		
Production/trade supervisors	**85.71**	Proprietor	Manager	503, 553-8, 613, 633, 803, 823, 843	600, 620,700,770,900,924
Office/clerical supervisors	**82.78**	Proprietor	Manager	303-7, 413-5, 433, 448, 456	370-3,400-1,420-1,430-2,470-1,500
Jurists	**73.08**	Proprietor	Manager	178-9	210-1
Farm/agricultural managers	**70.37**	Proprietor	Manager	475-7, 485	20
Health professionals	**69.72**	Proprietor	Manager	84-9, 96	300-1,304-6,312,325-6
Clergy	**69.70**	Proprietor	Manager	176	204-5
Execs, managers, administrators	**66.51**	Proprietor	Manager	3-22**,23-37	1-16,22-43,50-3,56,62-95
FIRE/sales professionals	**60.33**	Proprietor	Manager	243-55	481-2,492
Construction/extraction trades	47.66	Ind. Producer	Worker	563-99, 615-7	622-53,680-92
Architects, engineers	44.24	Ind. Producer	Worker	43-59	130-53
Farmers, agricultural workers	42.22	Ind. Producer	Worker	473-4, 479-84, 486-98	21,601-13
Writers, artists, entertainers	38.72	Ind. Producer	Worker	183-99	260-96
Social/recreational workers	38.71	Ind. Producer	Worker	174-5, 177	201-2,206
Scientists	34.77	Ind. Producer	Worker	63-83, 166-73	120-4,160-86
Mechanics	31.03	Ind. Producer	Worker	505-49	670,701-62
Sales workers	30.31	Ind. Producer	Worker	256-85	472-80,483-90,493-6
Technicians and programmers	30.06	Ind. Producer	Worker	213-35	100-11,154-6,190-6,214-5,903-4
Precision production workers	28.63	Ind. Producer	Worker	634-99	621,780-4,806-10,813,821,823,833-5,844-52,875-6,891-2
Other laborers	25.59	Ind. Producer	Worker	863-89	660,671-6,693-4,894-6,942-75
Service workers	24.44	Ind. Producer	Worker	403-7, 416-31, 434-47, 449-55, 457-69	374-95,402-16,422-25,434-65,905
Fabricators, inspectors	24.35	Ind. Producer	Worker	783-99	666,771-75,814,874,941
Clerical workers	24.05	Ind. Producer	Worker	308-89	54,60,501-94
Vehicle/machine operators	23.30	Ind. Producer	Worker	703-79, 804-13, 824-9, 844-59	785-804,812,815-20,822,824-32,836-43,853-73,880-90,893,911-23,926-36
Nurses, health techs	17.82	Ind. Producer	Worker	95, 97-106, 203-8	303,311,313-24,330-65
Teachers	15.18	Ind. Producer	Worker	113-65	200,220-55

Notes: The percentages of respondents in each occupational group that report having supervisory responsibilities come from a GSS sample that includes respondents who are 18 to 65 years old, work full-time, and have nonmissing occupation and supervisory data in the 1988 to 2010 waves. Bold font is used to highlight those occupations in which >60 percent of respondents report having supervisory responsibilities.

**These occupational categories define the upper stratum of the managerial class (i.e., high-level managers)

Table B2. Class distributions by Occupational Proxy Measure, GSS 1994-2010

Variable	Marginal Pct		
	Direct Measure	Occ. Proxy Measure	Alt. Occ. Proxy Measure
Class			
Workers	57.90	67.56	63.23
Ind. producers	5.76	6.67	5.24
Managers	30.60	20.90	25.28
Proprietors	5.73	4.88	6.24
Missclassification rate	-	34.08	34.53
Class strata			
Managers			
Low-level managers	20.63	11.39	15.72
High-level managers	9.97	9.50	9.56
Proprietors			
Small proprietors	4.11	3.47	4.71
Large proprietors	1.62	1.41	1.53
Missclassification rate	-	39.74	40.71

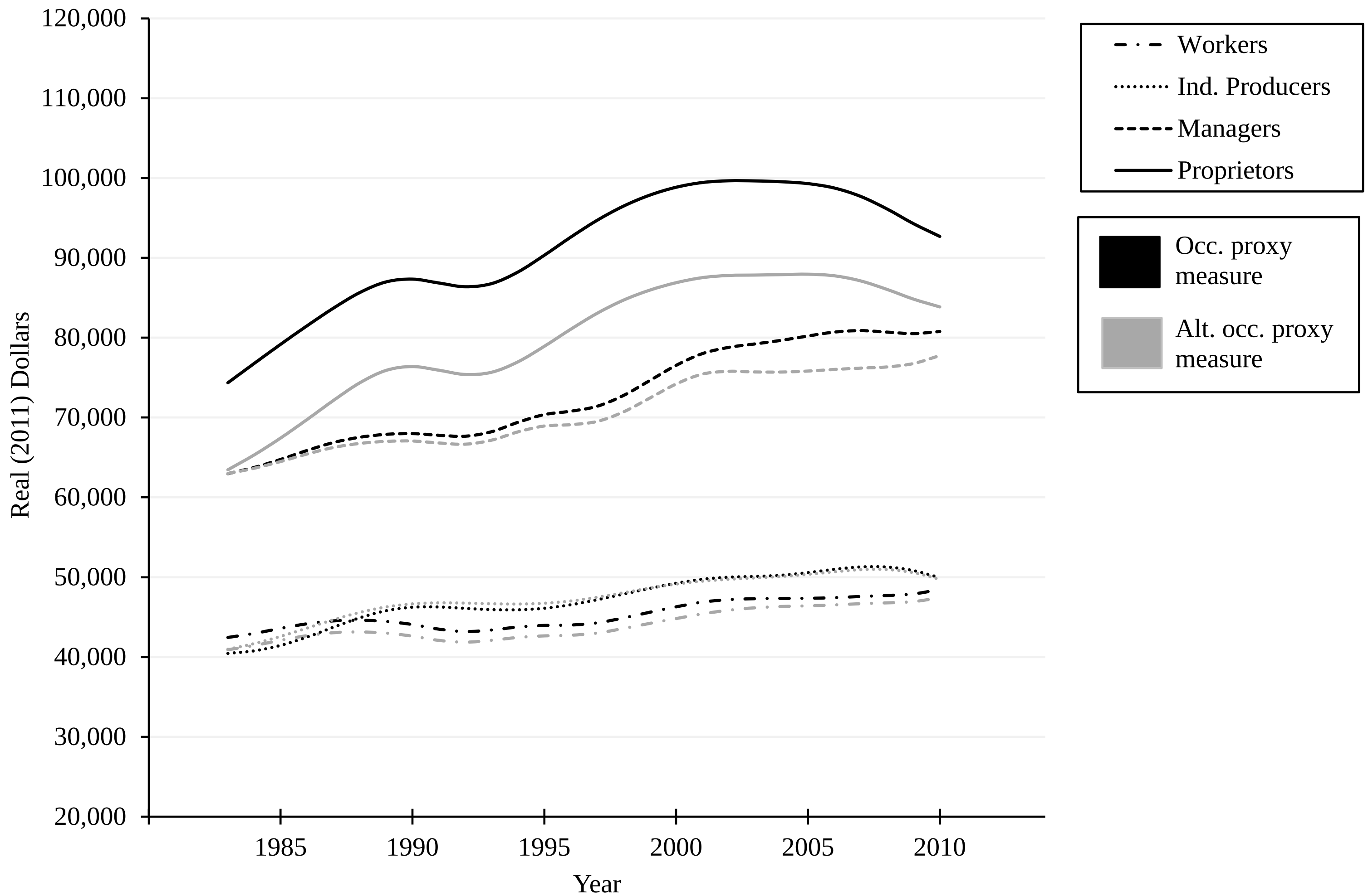
Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1994 to 2010 GSS waves. Results are combined estimates from 5 multiple imputation datasets. Missclassification rate gives the percentage of respondents classified differently under the direct versus occupational proxy measure. The alternate occupational proxy measure additionally classifies architects, engineers, production trades, extraction trades, and farmers as proprietors or managers (depending on their self-employment status), rather than independent producers or workers.

Table B3. Estimates and Fit Statistics from Static Models of Income by Occupational Proxy Measure, GSS 1994-2010

Specification	Unadjusted Income Model						Adjusted Income Model					
	Direct Measure		Occ. Proxy Measure		Alt. Occ. Proxy Measure		Direct Measure		Occ. Proxy Measure		Alt. Occ. Proxy Measure	
	coef	se	coef	se	coef	se	coef	se	coef	se	coef	se
Class												
Workers	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
Ind. producers	6107	2102	9736	2031	11383	2546	-2647	2014	2066	1883	3604	2374
Managers	16658	1139	24799	1338	23858	1248	11269	987	15807	1207	14186	1165
Proprietors	53920	4128	59791	3797	49832	3246	38415	3431	39960	3567	32000	3059
Rsq	0.069		0.089		0.081		0.261		0.263		0.256	
Class strata												
Workers	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
Ind. producers	6107	2102	9736	2031	11383	2546	-2473	2029	2140	1884	3692	2371
Managers												
Low-level managers	9521	1094	19500	1704	18801	1479	6923	1013	13079	1550	11208	1399
High-level managers	31433	2590	31153	2037	32175	2001	20733	2083	19332	1852	19523	1821
Proprietors												
Small proprietors	37704	3689	40762	3893	32718	3055	23837	3369	22697	3606	16564	2843
Large proprietors	95219	10962	106738	8403	102692	8140	77062	9168	83380	8098	80573	7830
Rsq	0.096		0.109		0.107		0.278		0.278		0.275	

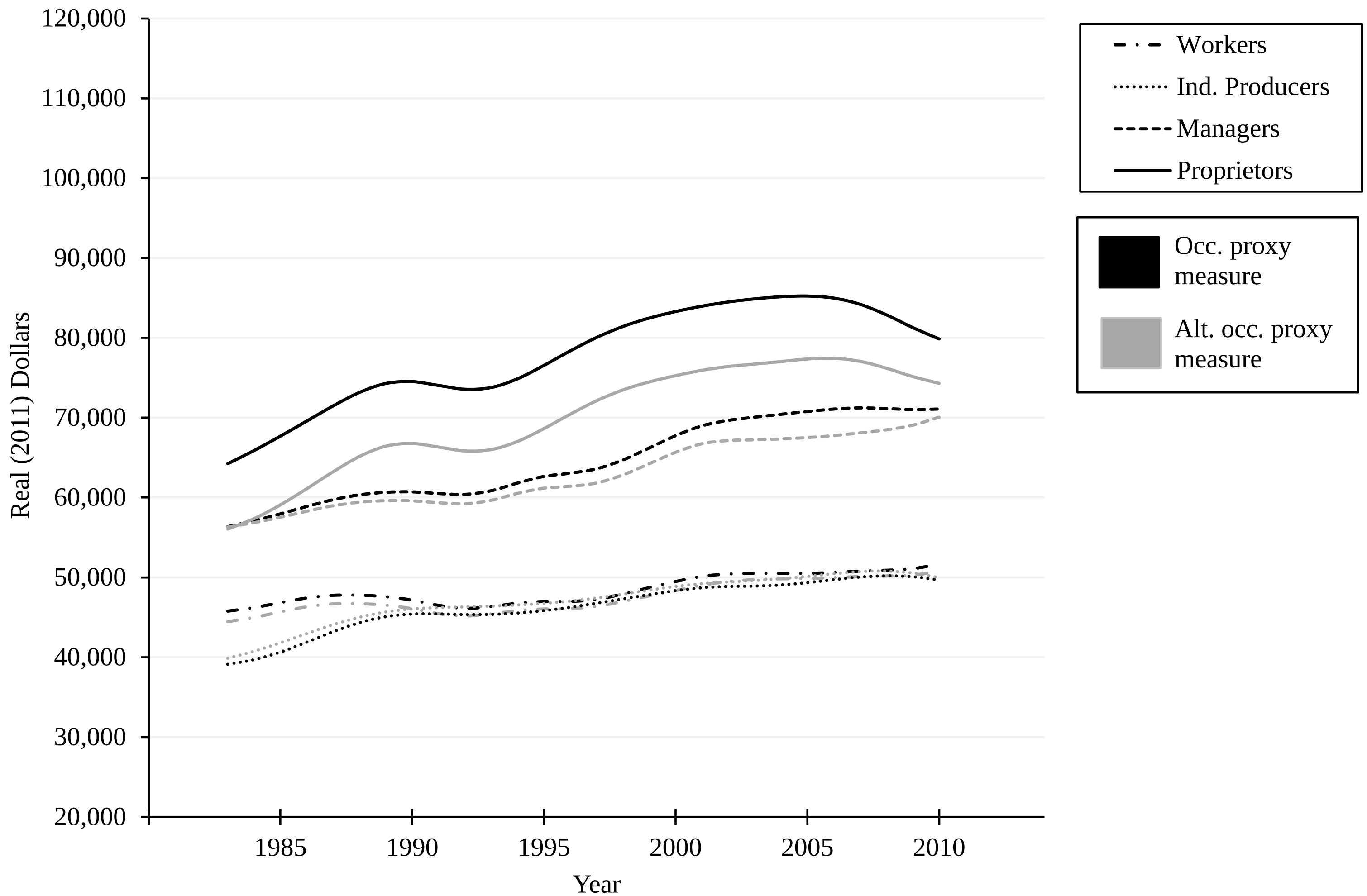
Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1994 to 2010 GSS waves. Results are combined estimates from 5 multiple imputation datasets. Heteroskedasticity-robust standard errors are reported. Adjusted models control for time, age, race, sex, education, region, cognitive ability, and parental education.

Figure B1. Unadjusted Trends in Mean Income by Class and Occupational Proxy Measure, CPS



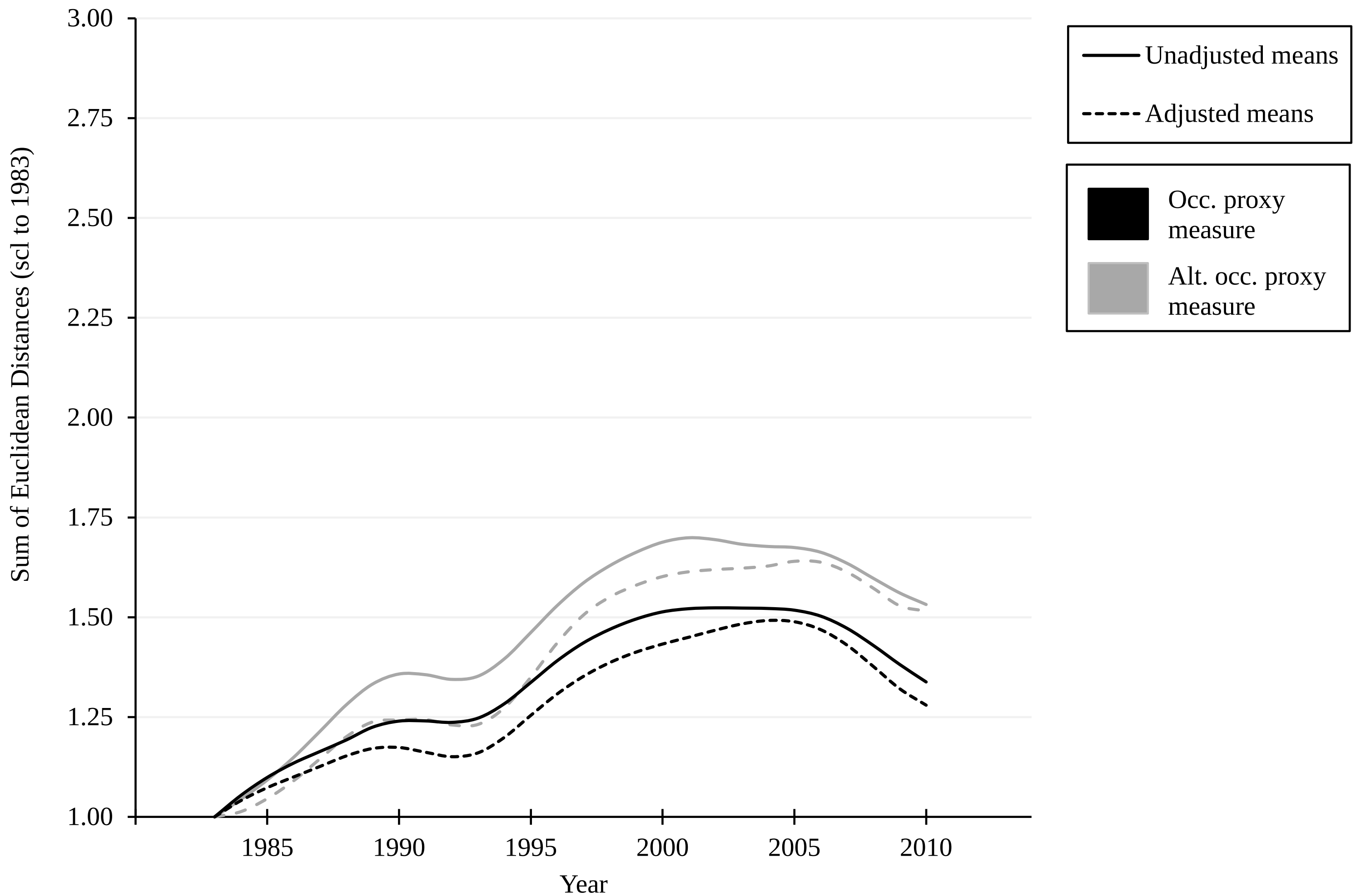
Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1983 to 2010 CPS waves. Results are combined estimates from 5 multiple imputation datasets.

Figure B2. Covariate-adjusted Trends in Mean Income by Class and Occupational Proxy Measure, CPS



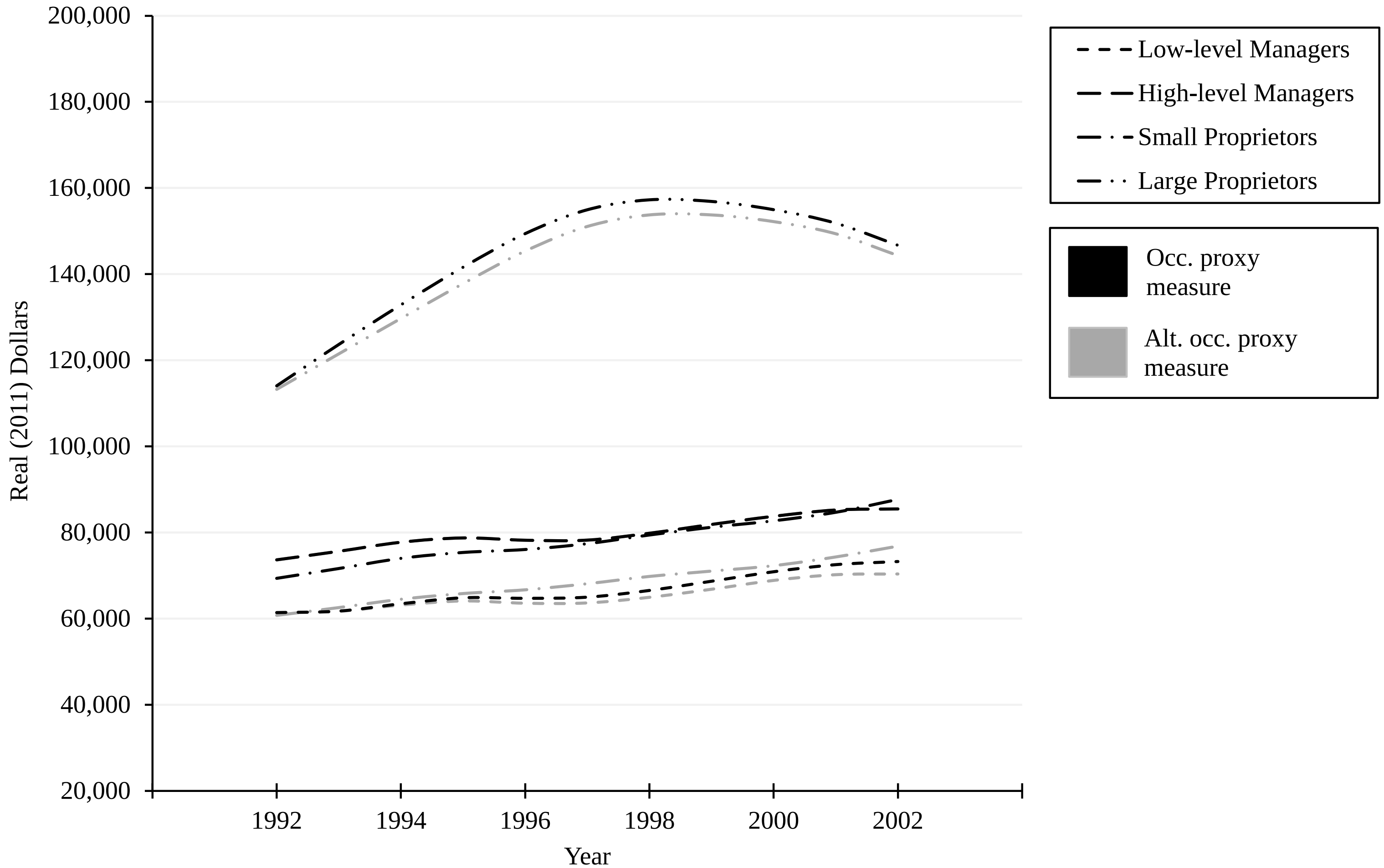
Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1983 to 2010 CPS waves. Results are combined estimates from 5 multiple imputation datasets.

Figure B3. Total Change in Mean Income Differences between Classes by Occupational Proxy Measure, CPS



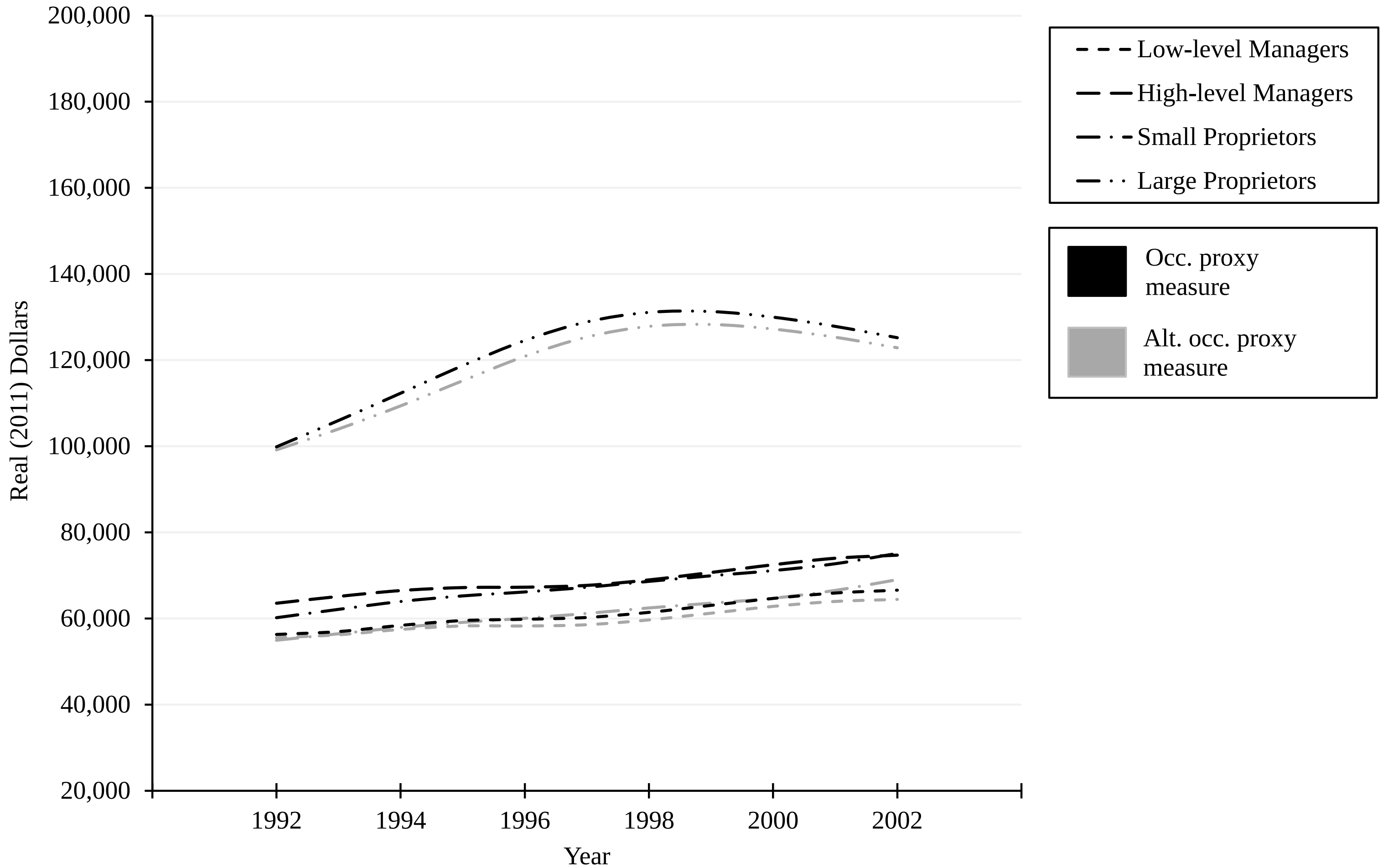
Notes: Samples include respondents who are 18 to 65 years old and work full-time in the 1983 to 2010 CPS waves. Results are combined estimates from 5 multiple imputation datasets.

Figure B4. Unadjusted Trends in Mean Income by Class Strata and Occupational Proxy Measure, CPS



Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1992-2002 CPS waves.
Results are combined estimates from 5 multiple imputation datasets.

Figure B5. Covariate-adjusted Trends in Mean Income by Class Strata and Occupational Proxy Measure, CPS



Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1992-2002 CPS waves.
Results are combined estimates from 5 multiple imputation datasets.

Part C: Income Measurement in the GSS and CPS

This section describes the procedures used to measure personal market income in the GSS and CPS, and it explains the adjustments applied to topcoded incomes in both surveys. In the GSS, total personal income earned over the previous year is measured in intervals, and dollar values are imputed based on interval midpoints. The highest incomes fall in a final open-ended interval, the lower bound of which increases systematically across survey ways as nominal incomes increase over time. Table C1 provides the exact lower-bound values used as topcode thresholds in the GSS as well as the percentage of respondents at each wave with incomes above these thresholds. Incomes are topcoded at $50,000 from 1983 to 1985, at $60,000 from 1986 to 1990, at $75,000 from 1991 to 1997, at $110,000 from 1998 to 2005, and at $150,000 from 2006 onward.

For these topcoded incomes, nominal values are estimated using a Pareto approximation for the upper tail of the income distribution (Hout 2004). This approach involves extrapolating from the next-to-last income interval's midpoint using the frequencies from both the next-to-last and the last open-ended intervals. Specifically, the formula for estimating topcoded incomes is

$$Y_{top} = LL_{top}\left(\frac{V}{V-1}\right), \tag{C1}$$

where LL_{top} is the lower bound of the last open-ended income interval and

$$V = \frac{\log(F_{top-1}+F_{top})-\log(F_{top})}{\log(LL_{top})-\log(LL_{top-1})}. \tag{C2}$$

In Equation C2, F_{top-1} and F_{top} are the frequencies from the next-to-last and the last open-ended intervals, respectively. These estimates are computed from data pooled across waves that use the same income measurement intervals because estimates computed separately for each survey wave are somewhat unstable with several extremely large imputed values. Table C1 reports the exact values substituted for topcoded incomes based on this imputation procedure.

In the CPS, income from different sources is measured in nominal dollars using separate survey items, and then these amounts are summed to achieve a measure of total market income. This study focuses on income from labor, businesses, farm operations, and several different types of investments, including income from interest, dividends, and rents. The survey items used to measure income from these different sources were revised between 1987 and 1988, so information on income measurement and topcoding is presented separately for the periods before and after this change. Between 1983 and 1987, the CPS used separate survey items to measure income from labor, businesses, farm operations, interest, and dividends and rents combined. From 1988 onward, the CPS first asked about income from a respondent's main job, regardless of whether that income came from labor, a business, or a farm. The survey then inquired about other income from secondary labor, business, or farm activities not reported as income from a respondent's main job. Additionally, after 1988, the CPS included separate items for income from dividends and rents.

In public CPS data, large incomes from each of these sources are top-coded to protect respondent anonymity. The exact topcoding thresholds for each income source and each survey wave, together with the percentage of topcoded respondents, are presented in Tables C2 to C4. To adjust for topcoding in the public CPS, incomes above the topcoding threshold are imputed with group-specific means of above-threshold incomes. These means are computed from internal CPS data (not available to the public) in which income topcoding is much less extensive. Internal CPS income data are not completely uncensored, but the internal topcoding thresholds are generally much higher than topcoding thresholds in the public data (e.g., from 2003 to 2010, the internal topcoding threshold for main job income was $1,099,999, while the public threshold was $200,000). Overall, internal CPS data contain uncensored incomes from all sources for about 99.5 to 99.8 percent of the sample (Larrimore et al 2008). Thus, internal CPS data can be used to estimate topcoded incomes for the public CPS while maintaining respondent anonymity, but because of internal topcoding, even these estimates will slightly understate average incomes for respondents above the public topcoding threshold. Nevertheless, this procedure provides more accurate estimates of topcoded incomes than other common adjustments, such as imputing based on a constant multiple of the public topcoding threshold (Card and DiNardo 2002; Larrimore et al 2008). The group means used to impute topcoded incomes in the public CPS are provided in Tables C5 to C9. From 1996 to 2010, the Census Bureau directly provided these group means along with the public CPS data. For survey waves prior to 1996, group means are obtained from Larrimore et al (2008), who calculated and reported this information with special permission from the Census Bureau in order to provide researchers with a consistent imputation procedure over time.

There are typically between 1 to 5 percent of respondents with incomes that exceed topcode thresholds at any given wave in both the GSS and CPS. For certain groups of respondents, however, a significantly higher proportion have topcoded incomes. In particular, about 8 percent of proprietors and 18 percent of large proprietors have incomes that exceed the topcode threshold for main job earnings in the CPS, and in the GSS, about 18 percent of proprietors and 25 percent of large proprietors have topcoded incomes. This indicates that income trends for proprietors and large proprietors, and especially trends in the upper quantiles of the income distribution for these groups, are sensitive to topcoding and should therefore be interpreted with caution. Furthermore, because the topcoding adjustments in this study do not reflect the enormous gains among the upper fractiles (i.e., the top 0.05 to 0.01 percent) of the income distribution (Piketty and Saez 2006) and because proprietors disproportionately have incomes in these fractiles, the estimates provided in this study likely understate the true increase in class income inequality since the early 1980s.

Table C1. Topcoded income data in the 1983-2010 GSS waves

Year	Income		
	Pct topcoded	Topcode threshold	Imputed value
1983	3.16	50,000	80,732
1984	4.28	50,000	80,732
1985	6.73	50,000	80,732
1986	3.26	60,000	93,729
1987	2.79	60,000	93,729
1988	3.14	60,000	93,729
1989	3.29	60,000	93,729
1990	3.81	60,000	93,729
1991	2.37	75,000	127,858
1993	5.74	75,000	127,858
1994	4.11	75,000	127,858
1996	5.30	75,000	127,858
1998	1.28	110,000	235,730
2000	1.71	110,000	235,730
2002	1.60	110,000	235,730
2004	2.88	110,000	235,730
2006	2.60	150,000	277,607
2008	3.29	150,000	277,607
2010	2.40	150,000	277,607

Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1983 to 2010 GSS waves. Imputed values come from a Pareto approximation of the upper tail of the income distribution.

Table C2. Topcoded income data in the 1983-1987 CPS waves

Year	Labor income		Business income		Farm income		Interest		Dividends/rents	
	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold
1983	0.92	75,000	0.21	75,000	0.02	75,000	0.00	75,000	0.02	75,000
1984	0.94	75,000	0.27	75,000	0.03	75,000	0.00	75,000	0.01	75,000
1985	0.51	99,999	0.16	99,999	0.00	99,999	0.02	99,999	0.03	99,999
1986	0.62	99,999	0.13	99,999	0.00	99,999	0.00	99,999	0.02	99,999
1987	0.83	99,999	0.17	99,999	0.01	99,999	0.01	99,999	0.01	99,999

Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1983 to 1987 CPS waves.

Table C3. Topcoded income data in the 1988-2010 CPS waves

Year	Main job income		Oth. labor income		Oth. business income		Oth. farm income	
	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold
1988	0.95	99,999	0.00	99,999	0.01	99,999	0.00	99,999
1989	1.19	99,999	0.00	99,999	0.01	99,999	0.00	99,999
1990	1.61	99,999	0.01	99,999	0.01	99,999	0.00	99,999
1991	1.56	99,999	0.00	99,999	0.03	99,999	0.00	99,999
1992	1.60	99,999	0.01	99,999	0.01	99,999	0.00	99,999
1993	1.83	99,999	0.00	99,999	0.01	99,999	0.00	99,999
1994	2.25	99,999	0.03	99,999	0.02	99,999	0.00	99,999
1995	2.61	99,999	0.04	99,999	0.02	99,999	0.01	99,999
1996	0.86	150,000	0.57	25,000	0.12	40,000	0.04	25,000
1997	1.04	150,000	0.42	25,000	0.10	40,000	0.01	25,000
1998	1.10	150,000	0.80	25,000	0.13	40,000	0.05	25,000
1999	1.12	150,000	0.73	25,000	0.15	40,000	0.03	25,000
2000	1.37	150,000	0.87	25,000	0.14	40,000	0.03	25,000
2001	1.57	150,000	1.03	25,000	0.12	40,000	0.08	25,000
2002	1.73	150,000	0.88	25,000	0.17	40,000	0.10	25,000
2003	1.04	200,000	0.42	35,000	0.10	50,000	0.07	25,000
2004	0.96	200,000	0.48	35,000	0.16	50,000	0.11	25,000
2005	0.96	200,000	0.44	35,000	0.14	50,000	0.08	25,000
2006	1.09	200,000	0.53	35,000	0.13	50,000	0.13	25,000
2007	1.20	200,000	0.61	35,000	0.13	50,000	0.06	25,000
2008	1.19	200,000	0.55	35,000	0.09	50,000	0.05	25,000
2009	1.44	200,000	0.61	35,000	0.11	50,000	0.05	25,000
2010	1.46	200,000	0.46	35,000	0.11	50,000	0.07	25,000

Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1988 to 2010 CPS waves.

Table C4. Topcoded income data in the 1988-2010 CPS waves continued

	Interest		Dividends		Rents	
Year	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold	Pct topcoded	Topcode threshold
1988	0.00	99,999	0.00	99,999	0.00	99,999
1989	0.02	99,999	0.00	99,999	0.01	99,999
1990	0.04	99,999	0.01	99,999	0.00	99,999
1991	0.02	99,999	0.01	99,999	0.00	99,999
1992	0.01	99,999	0.00	99,999	0.01	99,999
1993	0.02	99,999	0.00	99,999	0.01	99,999
1994	0.02	99,999	0.00	99,999	0.02	99,999
1995	0.01	99,999	0.01	99,999	0.02	99,999
1996	0.03	99,999	0.01	99,999	0.03	99,999
1997	0.03	99,999	0.02	99,999	0.02	99,999
1998	0.05	99,999	0.07	99,999	0.07	99,999
1999	0.21	35,000	0.59	15,000	0.23	25,000
2000	0.32	35,000	0.70	15,000	0.19	25,000
2001	0.24	35,000	0.59	15,000	0.19	25,000
2002	0.25	35,000	0.38	15,000	0.25	25,000
2003	0.23	25,000	0.23	15,000	0.15	40,000
2004	0.23	25,000	0.39	15,000	0.13	40,000
2005	0.36	25,000	0.43	15,000	0.16	40,000
2006	0.44	25,000	0.53	15,000	0.14	40,000
2007	0.60	25,000	0.51	15,000	0.17	40,000
2008	0.60	25,000	0.63	15,000	0.13	40,000
2009	0.40	25,000	0.40	15,000	0.17	40,000
2010	0.38	25,000	0.54	15,000	0.17	40,000

Notes: Sample includes respondents who are 18 to 65 years old and work full-time in the 1988 to 2010 CPS waves.

Table C5. Group mean imputation values for topcoded incomes in the 1983-1987 CPS waves

Year	Labor income						Business income					
	Male			Female			Male			Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
1983	89,485	87,647	96,915	92,340	87,647	NA	88,987	90,964	90,964	90,608	NA	90,964
1984	90,220	NA	92,530	88,528	NA	NA	92,506	86,400	86,400	89,529	NA	NA
1985	99,999[a]	99,999[a]	99,999[a]	99,999[a]	NA	NA	99,999[a]	NA	99,999[a]	99,999[a]	NA	NA
1986	136,613	170,804	124,324	133,348	NA	NA	136,144	108,836	NA	108,836	NA	108,836
1987	140,359	119,934	150,042	125,434	169,047	NA	130,751	-99,999	170,968	170,968	NA	NA

Notes: NA indicates that there were not any topcoded respondents in a particular group.
[a]Public topcode threshold is equivalent to internal threshold and thus no additional information about large incomes is available.

Table C6. Group mean imputation values for topcoded incomes in the 1983-1987 CPS waves continued

Year	Farm income						Interest	Dividends/rents
	Male			Female				
	White	Black	Hispanic	White	Black	Hispanic		
1983	82,381	NA	NA	NA	NA	NA	97,565	92,724
1984	83,154	NA	NA	NA	NA	NA	94,024	87,201
1985	99,999[a]	NA	NA	NA	NA	NA	99,999[a]	99,999[a]
1986	NA	NA	NA	NA	NA	NA	99,999[a]	99,999[a]
1987	122,398	NA	NA	NA	NA	NA	99,999[a]	99,999[a]

Notes: NA indicates that there were not any topcoded respondents in a particular group.
[a]Public topcode threshold is equivalent to internal threshold and thus no additional information about large incomes is available.

Table C7. Group mean imputation values for topcoded incomes in the 1988-2010 CPS waves

Year	Main job income						Other labor income					
	Male			Female			Male			Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
1988	148,852	136,582	159,300	151,838	153,098	153,098	NA	NA	NA	NA	NA	NA
1989	143,204	138,971	154,412	152,647	137,250	137,250	NA	NA	NA	NA	NA	NA
1990	153,067	159,309	153,072	143,812	124,782	124,782	99,999[a]	NA	NA	NA	NA	NA
1991	151,763	144,161	135,010	153,090	132,453	132,230	NA	NA	NA	NA	NA	NA
1992	142,991	133,707	136,560	131,061	121,099	NA	99,999[a]	NA	NA	NA	NA	NA
1993	148,241	144,800	143,657	149,557	114,123	114,123	99,999[a]	NA	NA	NA	NA	NA
1994	188,027	232,995	205,449	215,571	273,701	159,042	158,174	125,569	NA	125,569	NA	NA
1995	187,347	180,854	179,894	191,029	160,143	212,792	207,148	109,775	109,775	109,775	NA	NA
1996	302,539	464,782	257,390	283,525	NA	404,570	64,542	29,778	183,748	56,977	35,662	35,662
1997	318,982	391,163	384,160	357,884	454,816	454,816	45,749	62,044	62,044	48,635	257,102	62,044
1998	330,659	204,325	309,950	306,468	267,659	394,555	61,345	51,707	39,943	48,755	47,530	35,080
1999	306,731	266,303	419,044	402,204	492,657	367,181	59,925	51,139	52,678	35,583	34,826	36,826
2000	300,974	257,525	362,315	256,384	244,810	333,565	50,037	35,625	39,676	51,469	67,776	50,770
2001	329,998	277,959	345,181	318,370	247,864	469,588	55,436	63,610	41,197	44,451	35,358	36,064
2002	320,718	326,969	331,926	361,315	477,562	330,981	60,670	49,155	50,534	43,389	40,556	65,493
2003	390,823	443,501	562,913	480,607	336,975	595,494	91,360	60,724	49,866	55,255	48,548	57,290
2004	404,469	360,083	427,646	390,847	556,932	387,963	89,988	156,017	64,536	67,710	57,293	49,202
2005	422,850	471,917	421,411	474,404	713,263	366,935	77,282	60,016	57,768	63,911	53,189	60,586
2006	423,545	543,488	404,840	410,175	303,536	257,855	79,378	52,371	54,590	55,344	59,002	68,283
2007	437,528	579,599	619,221	423,652	615,203	438,937	74,091	53,197	56,306	61,472	45,266	240,674
2008	419,969	272,589	425,629	366,022	688,117	351,023	73,029	51,636	56,732	57,459	57,069	57,991
2009	389,599	553,087	436,465	407,720	383,596	397,063	72,946	62,071	57,838	59,069	70,577	64,367
2010	409,068	418,365	415,929	433,605	566,972	493,804	67,527	59,900	54,534	62,705	40,348	87,422

Notes: NA indicates that there were not any topcoded respondents in a particular group.
[a]Public topcode threshold is equivalent to internal threshold and thus no additional information about large incomes is available.

Table C8. Group mean imputation values for topcoded incomes in the 1988-2010 CPS waves continued

Year	Other business income						Other farm income					
	Male			Female			Male			Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
1988	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1989	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1990	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1991	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1992	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1993	99,999[a]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1994	157,513	NA	NA	157,513	NA	NA	NA	NA	NA	NA	NA	NA
1995	305,001	NA	357,471	357,471	NA	NA	NA	NA	NA	NA	NA	NA
1996	154,528	82,233	NA	64,058	NA	NA	53,067	45,717	NA	45,717	NA	NA
1997	128,477	152,709	152,709	152,709	NA	152,709	38,781	NA	NA	38,781	NA	NA
1998	101,769	NA	104,335	53,484	NA	NA	90,170	NA	NA	61,127	NA	NA
1999	123,543	NA	103,545	52,835	NA	NA	65,337	NA	NA	44,558	NA	NA
2000	119,583	NA	64,058	63,258	NA	NA	87,162	NA	51,354	54,785	NA	NA
2001	117,099	NA	55,231	55,231	NA	NA	NA	NA	NA	NA	NA	NA
2002	127,597	108,083	79,683	56,934	49,520	49,520	NA	NA	NA	NA	NA	NA
2003	141,605	149,560	149,560	75,880	NA	NA	65,682	199,326	199,326	46,839	NA	199,326
2004	111,645	104,785	91,233	99,464	104,785	NA	67,545	37,503	50,171	40,178	37,503	NA
2005	160,832	164,370	164,370	90,574	NA	164,370	50,413	45,639	45,639	71,842	45,639	45,639
2006	186,628	76,650	76,650	80,515	76,650	76,650	208,516	43,931	141,398	51,655	43,931	43,931
2007	125,974	195,957	146,920	454,133	195,957	NA	145,701	48,679	48,679	43,376	NA	48,679
2008	93,400	93,014	93,014	93,014	93,014	93,014	43,286	NA	49,095	63,438	NA	49,095
2009	133,732	183,401	183,401	171,377	183,401	NA	62,368	NA	35,615	NA	NA	NA
2010	120,846	135,215	135,215	122,432	NA	135,215	61,670	59,646	NA	34,352	NA	NA

Notes: NA indicates that there were not any topcoded respondents in a particular group.
[a]Public topcode threshold is equivalent to internal threshold and thus no additional information about large incomes is available.

Table C9. Group mean imputation values for topcoded incomes in the 1988-2010 CPS waves continued

Year	Interest`	Dividends	Rents
1988	99,999[a]	99,999[a]	99,999[a]
1989	99,999[a]	99,999[a]	99,999[a]
1990	99,999[a]	99,999[a]	99,999[a]
1991	99,999[a]	99,999[a]	99,999[a]
1992	99,999[a]	99,999[a]	99,999[a]
1993	99,999[a]	99,999[a]	99,999[a]
1994	99,999[a]	99,999[a]	99,999[a]
1995	99,999[a]	99,999[a]	99,999[a]
1996	99,999[a]	99,999[a]	99,999[a]
1997	99,999[a]	99,999[a]	99,999[a]
1998	99,999[a]	99,999[a]	99,999[a]
1999	60,819	36,877	57,453
2000	63,005	36,962	55,220
2001	61,337	36,364	58,676
2002	64,854	38,962	57,417
2003	50,186	33,581	72,409
2004	51,372	39,987	74,636
2005	55,524	35,416	76,259
2006	54,984	37,508	76,212
2007	53,946	38,224	75,061
2008	52,619	33,651	70,556
2009	51,580	38,815	73,177
2010	55,289	40,100	71,580

[a]Public topcode threshold is equivalent to internal threshold and thus no additional information about large incomes is available.

Part D: IP-weighted Local Linear Regression

This section describes IP weighting and local linear regression with a series of simulation experiments. Consider first a simple example in which the goal is to estimate temporal trends in the mean of a metric outcome separately for two distinct groups without any compositional differences between them. The nonparametric regression function in this example can be written as

$$E(Y|C = j,t) = m_j(t), \quad j = 1,2 \tag{D1}$$

where $E(Y|C = j,t)$ is the expectation of the outcome for group $C = j$ at time t, and $m_j(t)$ is a group-specific function of time. The conventional local linear estimator for $m_j(t)$ is given by

$$\widehat{m}_j(t) = \sum_{i=1}^{n_j} a_{ij}y_{ij} = \sum_{i=1}^{n_j} \left(\frac{1}{n_j}\right)\frac{\left(S_2(t)-S_1(t)(t_{ij}-t)\right)K_b(t_{ij}-t)}{S_2(t)S_0(t)-S_1(t)^2}y_{ij}. \tag{D2}$$

In this equation, y_{ij} is the metric outcome for individual i in group j, and

$a_{ij} = \left(\frac{1}{n_j}\right)\frac{\left(S_2(t)-S_1(t)(t_{ij}-t)\right)K_b(t_{ij}-t)}{S_2(t)S_0(t)-S_1(t)^2}$ is a local weight assigned to each respondent, where

$S_v(t) = \sum_{i=1}^{n_j}\left(\frac{1}{n_j}\right)\left(t_{ij}-t\right)^v K_b\left(t_{ij}-t\right)$ and $K_b\left(t_{ij}-t\right) = \frac{1}{b}K\left(\frac{t_{ij}-t}{b}\right)$ represents a kernel function with bandwidth b.

The asymptotic bias of $\widehat{m}_j(t)$ is equal to

$$Bias\left(\widehat{m}_j(t), m_j(t)\right) = \frac{b^2}{2}m_j''(t)\int u^2K_b(u)du. \tag{D3}$$

This expression indicates that the bias of the conventional local linear regression estimator depends on the size of the bandwidth, the curvature of the conditional expectation function, and the type of kernel. Specifically, smaller bandwidths and less curvature in the expectation function are associated with lower bias because as $b \to 0$ or $m_j''(t) \to 0$, $Bias\left(\widehat{m}_j(t), m_j(t)\right) \to 0$. When the conditional expectation function is linear, and thus $m_j''(t) = 0$, the local linear estimator is unbiased.

The asymptotic variance of $\widehat{m}_j(t)$ is given by

$$Var\left(\widehat{m}_j(t)\right) = \frac{\sigma^2(t)}{nbf(t)}\int uK_b(u)^2du, \tag{D4}$$

indicating that the precision of the estimator depends on the sample size, bandwidth, type of kernel, and the distribution of observations across time. Specifically, the variance decreases as the sample size, bandwidth, and density of observations at time t increase. Based on these expressions for the asymptotic bias and variance, the asymptotic mean squared error of $\widehat{m}_j(t)$ is equal to

$$AMSE\left(\widehat{m}_j(t)\right) = \frac{\sigma^2(t)}{nhf(t)} \int uK_b(u)^2 du + \left(\frac{b^2}{2} m_j''(t) \int u^2 K_b(u) du\right)^2. \quad \text{(D5)}$$

This reveals a clear bias-variance tradeoff associated with the size of the bandwidth: narrow bandwidths reduce bias but increase variance, and wide bandwidths reduce variance but increase bias. In general, narrow bandwidths are expected to perform best when the group-specific time trends are highly nonlinear, while wider bandwidths are better suited for conditional expectation functions that are characterized by more moderate deviations from linearity.

Now suppose that, rather than being comparable in terms of other covariates, the two groups of interest in this example differ on one or more factors that also affect the outcome. In this situation, the group-specific trends estimated from Equation D2 are confounded by compositional differences between groups. There are several approaches to adjusting trend estimates for this type of confounding. First, one could condition on the covariates, x, in a parametric model for $E(Y|C = j, X = x, t)$, but correctly specifying this model may be difficult if its functional form is unknown or highly nonlinear. The conditional expectation $E(Y|C = j, X = x, t)$ could also be estimated nonparametrically using multivariate local linear regression and a higher-order kernel, but with many covariates, this approach suffers from the so-called "curse of dimensionality," meaning that the variance of this estimator rapidly increases with the dimension of x. Even with a large sample and just a few covariates, multivariate local linear regression can perform quite poorly.

IP-weighting provides an alternative semi-parametric approach to adjusting group-specific time trends for confounding that attenuates some of the problems associated with parametric multivariate regression and fully nonparametric multivariate methods. With this approach, information about compositional differences between the groups of interest on a potentially high-dimensional set of covariates is reduced to a single dimension—the conditional probability of group membership given the covariates—and a simple transformation of this probability yields a set of weights that balance the distribution of covariates across groups. Conventional local linear regression is then applied to the weighted observations to estimate unconfounded group-specific time trends. Specifically, this approach estimates the following double expectation function: $E_x\big(E(Y|C=j,X=x,t)\big)=g_j(t)$, which is the expectation of the outcome conditional on group membership, covariates, and time, averaged across the covariates.

The IP-weighted local linear regression estimator for this function can be expressed as

$$\hat{g}_j(t)=\sum_{i=1}^{n_j} a_{ij}^{w} y_{ij}=\sum_{i=1}^{n_j}\left(\frac{w_{ij}}{\sum_{i=1}^{n_j} w_{ij}}\right)\frac{\big(S_2^w(t)-S_1^w(t)(t_{ij}-t)\big)K_b(t_{ij}-t)}{S_2^w(t)S_0^w(t)-S_1^w(t)^2}y_{ij}, \tag{D6}$$

where w_{ij} is the IP weight equal to $\frac{P(C=j|t_i)}{P(C=j|X=x_i,t_i)}$, $S_v^w(t)$ is equal to $\sum_{i=1}^{n_j}\left(\frac{w_{ij}}{\sum_{i=1}^{n_j} w_{ij}}\right)\big(t_{ij}-t\big)^v K_b\big(t_{ij}-t\big)$, and $K_b\big(t_{ij}-t\big)$ is defined as previously. The weight w_{ij} varies around 1 based on the degree to which covariates impact the probability of group membership. It gives greater weight to individuals who are less likely to be members of their observed group given their covariate values, and it gives less weight to individuals who are more likely to be a member of their observed group. This estimator is considered semi-parametric in the sense that the weights are in practice typically estimated from a parametric model, but this need not be the case, as fully nonparametric methods could also be used to estimate the weights. In the next section, I demonstrate how the IP-weighted local linear estimator works in several different simulated data examples, and then I investigate the small and large sample properties of this estimator with a series of simulation experiments.

Table D1 summarizes the distribution of a binary covariate by group membership and time from fifty thousand simulated observations. The exact data generating model is documented in the footnotes to the table (as is the case for all subsequent simulations). In this example, selection into different groups based on the binary covariate is invariant over time, and at each time point, this covariate is highly imbalanced across the groups of interest. The second set of columns in Table D1 illustrates the impact of weighting by the inverse probability of group membership, revealing covariate balance across groups at each time point in the weighted sample.

Figure D1 presents unadjusted and IP-weighted local linear regression estimates for a simulated outcome based on the group and covariate data from Table D1. The outcome variable is generated from a model with a highly nonlinear time trend. There are no differences in this trend between groups conditional on the binary covariate, which has time-invariant effects on both the probability of group membership and the outcome. Because the binary covariate is imbalanced across groups and has a time-invariant effect on the outcome, unadjusted estimates show a similar temporal trend for each group, but the trend lines are shifted upward and downward by a constant amount. The IP-weighted estimates, by contrast, closely approximate the true expectation function that both groups share conditional on the confounding covariate.

Figure D2 presents local linear regression estimates from simulated data where, as in the previous example, the outcome follows a highly nonlinear trend among both groups conditional on the binary covariate, and the binary covariate has a time-invariant effect on the probability group membership. Unlike the previous example, however, the binary covariate has time-dependent effects on the outcome, with its impact increasing in magnitude over time. For this example, unadjusted local linear estimates show highly divergent temporal trends between groups. The IP-weighted estimates, on the other hand, show both groups following the same nonlinear trend over time and closely approximate the true conditional expectation function averaged across levels of the binary covariate.

Table D2 and Figure D3 present results from a final simulated example in which the outcome follows a highly nonlinear trend, and the two groups of interest follow the same trend conditional on a confounding binary covariate. The binary covariate in this example has a time-invariant effect on the outcome and time-dependent effects on the probability of group membership such that compositional differences between groups on this covariate become more pronounced over time. The descriptive statistics in Table D2 summarize this pattern of growing covariate imbalance over time and show that IP weighting balances the covariate distribution across groups at each time point. Figure D3 plots unadjusted local linear estimates, which show divergent trends in the outcome between groups, and IP-weighted local linear estimates, which recover the time trend shared by both groups conditional on the binary covariate.

The simulated examples discussed here demonstrate the various confounding processes that can lead to divergent group-specific trend estimates with conventional local linear regression. In empirical research on class differences in income over time, all of these confounding processes are likely operating simultaneously with multiple covariates. For example, the effects of both education and gender on income are changing over time as are their effects on class attainment. The simulated examples discussed previously show that IP-weighted local linear regression is capable of adjusting for all of these confounding processes.

Table D3 and Table D4 describe the small and large sample properties of the IP-weighted local linear regression estimator based on the results of several simulation experiments with one thousand replications. The first simulation, summarized in the upper panels of Table D3 and Table D4, involves a highly nonlinear trend with a conditionally normal and homoscedastic outcome. The second simulation, summarized in the middle panels of Table D3 and Table D4, involves only a moderate degree of trend nonlinearity with a conditionally normal and homoscedastic outcome. The last simulation, presented in the lower panels of Table D3 and Table D4, is designed to approximate the income models considered in the empirical analysis. It involves a moderately nonlinear temporal trend with a highly skewed and heteroscedastic outcome.

In each of these simulations, the IP-weighted local linear regression estimator performs as expected. The bootstrap standard error provides a good approximation for the true standard deviation of IP-weighted estimates. Bias is more severe at a point on the expectation function with a high degree of curvature, compared with a point where the expectation function is roughly linear, and it increases with the size of the bandwidth. Shapiro-Wilk tests applied to the simulated sampling distributions suggest that estimates based on conditionally normal data are themselves normally distributed. For small sample estimates based on highly skewed lognormal data, these tests frequently reject the null of a normal sampling distribution. Graphical inspection of these distributions, however, indicates that they are nevertheless very close to normal. In simulations where temporal change in the outcome is only moderately nonlinear, bias with both the narrow and standard bandwidths is close to zero, and 95 percent bootstrap confidence intervals that assume a normal sampling distribution cover the true conditional expectation with equivalent probability. Overall, for moderately nonlinear functions, the IP-weighted local linear regression estimator performs quite well regardless of bandwidth size, and for highly nonlinear functions, the IP-weighted estimator performs better with a narrow bandwidth.

Table D1. Weighted and unweighted covariate distribution by group in simulated data example with time-invariant selection process[a]

Time	Unweighted		IP-weighted	
	Group 1	Group 2	Group 1	Group 2
1.00-1.49	0.29	0.73	0.50	0.51
1.50-2.49	0.31	0.72	0.53	0.51
2.50-3.49	0.31	0.70	0.52	0.49
3.50-4.49	0.27	0.70	0.47	0.50
4.50-5.49	0.28	0.70	0.48	0.49
5.50-6.49	0.29	0.69	0.49	0.49
6.50-7.49	0.30	0.69	0.51	0.49
7.50-8.49	0.29	0.69	0.49	0.49
8.50-9.49	0.29	0.71	0.49	0.50
9.50-10.49	0.31	0.71	0.52	0.51
10.50-11.49	0.29	0.70	0.48	0.49
11.50-12.49	0.29	0.70	0.49	0.50
12.50-13.49	0.31	0.69	0.51	0.49
13.50-14.49	0.28	0.69	0.47	0.49
14.50-15.49	0.30	0.71	0.49	0.50
15.50-16.49	0.30	0.68	0.50	0.48
16.50-17.49	0.29	0.71	0.48	0.51
17.50-18.49	0.31	0.72	0.52	0.51
18.50-19.49	0.31	0.70	0.51	0.50
19.50-20.00	0.27	0.74	0.45	0.53

[a]50,000 observations simulated from the following data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.4*covariate)

Table D2. Weighted and unweighted covariate distribution by group in simulated data example with time-dependent selection process[a]

Time	Unweighted		IP-weighted	
	Group 1	Group 2	Group 1	Group 2
1.00-1.49	0.48	0.49	0.49	0.46
1.50-2.49	0.48	0.54	0.50	0.50
2.50-3.49	0.49	0.56	0.51	0.51
3.50-4.49	0.49	0.56	0.52	0.50
4.50-5.49	0.45	0.57	0.48	0.50
5.50-6.49	0.44	0.61	0.49	0.52
6.50-7.49	0.44	0.61	0.49	0.51
7.50-8.49	0.43	0.58	0.50	0.48
8.50-9.49	0.45	0.63	0.53	0.51
9.50-10.49	0.42	0.62	0.50	0.50
10.50-11.49	0.43	0.65	0.53	0.51
11.50-12.49	0.41	0.64	0.51	0.50
12.50-13.49	0.38	0.65	0.49	0.50
13.50-14.49	0.36	0.68	0.48	0.52
14.50-15.49	0.34	0.64	0.47	0.48
15.50-16.49	0.36	0.66	0.50	0.49
16.50-17.49	0.35	0.67	0.51	0.49
17.50-18.49	0.35	0.71	0.53	0.52
18.50-19.49	0.30	0.71	0.48	0.51
19.50-20.00	0.31	0.68	0.52	0.49

[a]50,000 observations simulated from the following data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time)

Table D3. Small sample (n=500) properties of IP-weighted local linear estimator in three simulation experiments

Simulation description	Mean SD	Mean BSE	Linear bias	Curvature bias	RMISE	95% CI coverage	Distn shape
Normal outcome w/ highly nonlinear temporal change[a]							
Narrow bandwidth	3316	3267	-951	-2292	4172	0.89	normal
Standard bandwidth	2455	2452	-2026	-5459	4701	0.59	normal
Wide bandwidth	1905	1909	-2462	-7007	5123	0.40	normal
Normal outcome w/ moderately nonlinear temporal change[b]							
Narrow bandwidth	3381	3339	-77	-168	3689	0.94	normal
Standard bandwidth	2423	2429	-330	-406	2690	0.95	normal
Wide bandwidth	1840	1851	-672	-1460	2308	0.88	normal
Lognormal outcome w/ moderately nonlinear temporal change[c]							
Narrow bandwidth	5755	5570	-443	-535	5993	0.93	aprx norm
Standard bandwidth	4197	4130	-582	-543	4413	0.94	aprx norm
Wide bandwidth	3255	3246	-809	-1476	3581	0.92	aprx norm

Notes: Narrow, standard, and wide bandwiths respectively equal 0.5, 1, and 2 times h=1.06*sd(x)*(n^(-1/5)). Linear bias refers to bias at a point where the regression function is roughly linear, and curvature bias refers to bias at a point where the regression function is highly nonlinear. RMISE denotes root mean integrated squared error of the estimator.

[a]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ normal(40000 + 7000*cos(time) + group*(0 + 0*g(time)) covariate*(0 + 3000*time),10000)

[b]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ normal(40000 + 1000*time - 100*time^2 + group*(0 + 0*time + 0*time^2) + covariate*(10000 + 3000*time - 50*time^2),10000)

[c]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ covar*(10000+5000*time-125*time^2) + lognormal(10.6 + 0.015*time - 0.0015*time^2 + group*(0 + 0*time + 0*time^2),0.5)

Table D4. Large sample (n=5000) properties of IP-weighted local linear estimator in three simulation experiments

Simulation description	Mean SD	Mean BSE	Linear bias	Curvature bias	RMISE	95% CI coverage	Distn shape
Normal outcome w/ highly nonlinear temporal change[a]							
Narrow bandwidth	1265	1258	-441	-1058	1633	0.88	normal
Standard bandwidth	917	917	-1260	-3322	2588	0.39	normal
Wide bandwidth	695	698	-2347	-6312	4338	0.19	normal
Normal outcome w/ moderately nonlinear temporal change[b]							
Narrow bandwidth	1349	1337	-73	-115	1435	0.95	normal
Standard bandwidth	960	956	-211	-223	1047	0.95	normal
Wide bandwidth	699	699	-544	-755	961	0.83	normal
Lognormal outcome w/ moderately nonlinear temporal change[c]							
Narrow bandwidth	2220	2194	-81	-139	2288	0.95	normal
Standard bandwidth	1598	1590	-248	-226	1668	0.95	normal
Wide bandwidth	1188	1187	-604	-725	1371	0.91	normal

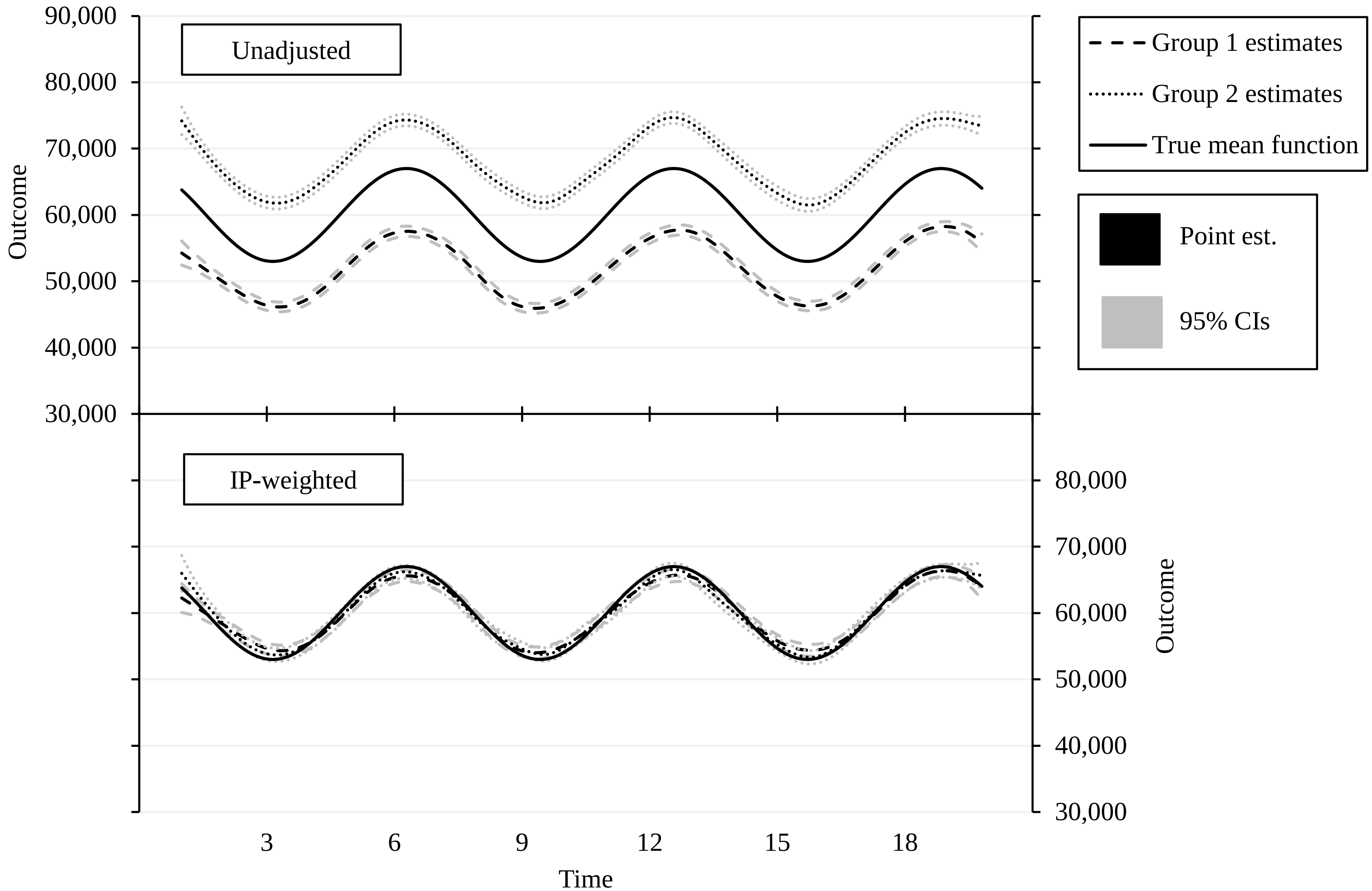
Notes: Narrow, standard, and wide bandwiths respectively equal 0.5, 1, and 2 times h=1.06*sd(x)*(n^(-1/5)). Linear bias refers to bias at a point where the regression function is roughly linear, and curvature bias refers to bias at a point where the regression function is highly nonlinear. RMISE denotes root mean integrated squared error of the estimator.

[a]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ normal(40000 + 7000*cos(time) + group*(0 + 0*g(time)) covariate*(0 + 3000*time),10000)

[b]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ normal(40000 + 1000*time - 100*time^2 + group*(0 + 0*time + 0*time^2) + covariate*(10000 + 3000*time - 50*time^2),10000)

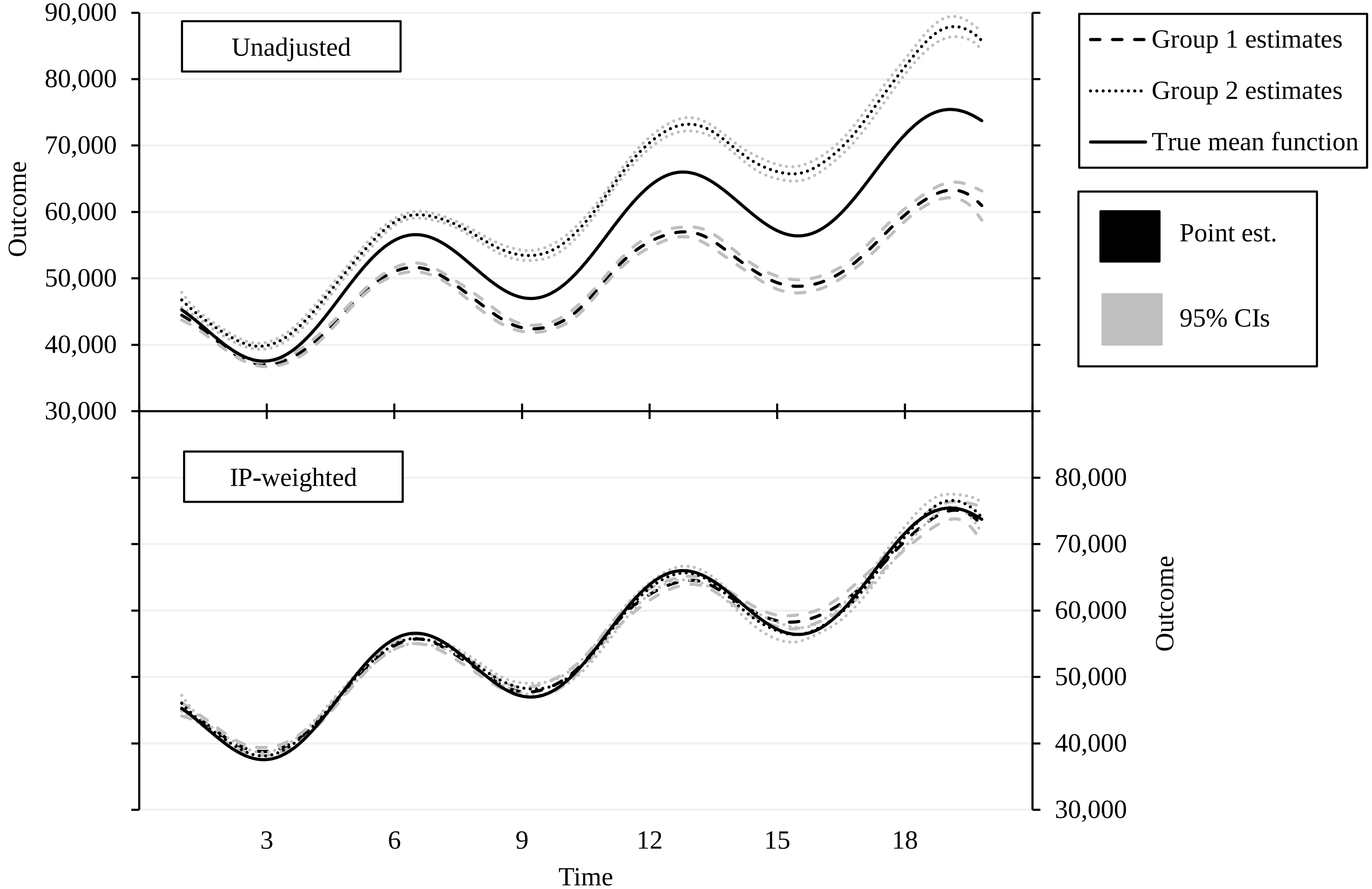
[c]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ covar*(10000+5000*time-125*time^2) + lognormal(10.6 + 0.015*time - 0.0015*time^2 + group*(0 + 0*time + 0*time^2),0.5)

Figure D1. Unadjusted and IP-weighted local linear estimates based on simulated data with time-invariant group selection and time-invariant covariate effects on outcome[a]



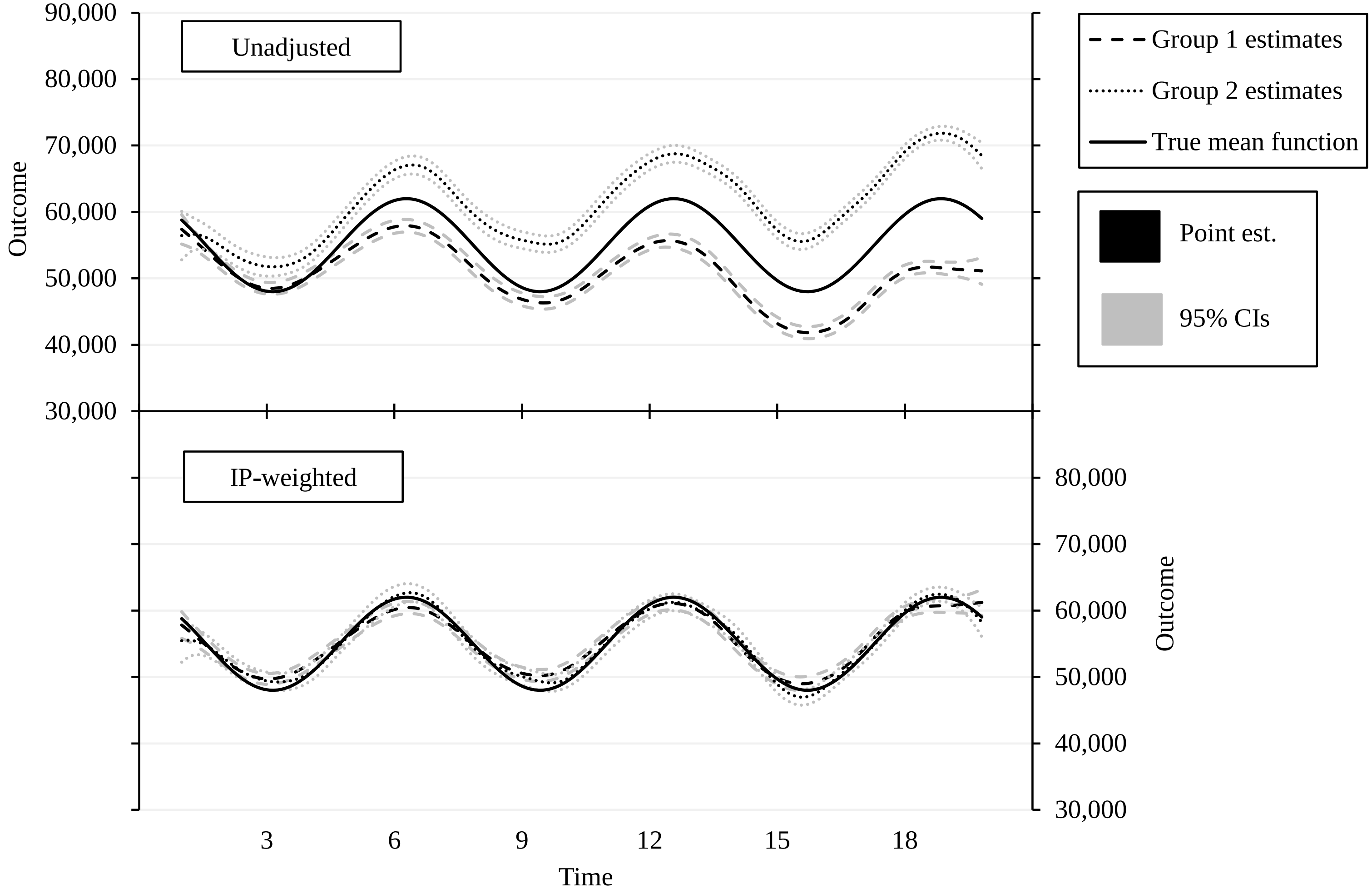
[a]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.4*covariate), outcome ~ normal(40000 + 7000*cos(time) + group*(0 + 0*g(time)) covariate*(40000 + 0*g(time)),10000)

Figure D2. Unadjusted and IP-weighted local linear estimates based on simulated data with time-invariant group selection and time-dependent covariate effects on outcome[a]



[a]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.4*covariate), outcome ~ normal(40000 + 7000*cos(time) + group*(0 + 0*g(time)) covariate*(0 + 3000*time),10000)

Figure D3. Unadjusted and IP-weighted local linear estimates based on simulated data with time-dependent group selection and time-invariant covariate effects on outcome[a]



[a]Data generating model: time ~ uniform(1,20), covariate ~ binomial(1,0.5), group ~ binomial(1,0.3 + 0.02*covariate*time), outcome ~ normal(40000 + 7000*cos(time) + group*(0 + 0*g(time)) covariate*(40000 + 0*g(time)),10000)