DIVING DEEPER: UNDERSTANDING DISPARITIES BETWEEEN BLACK AND WHITE RESIDENTS

MetroStats Technical Appendix

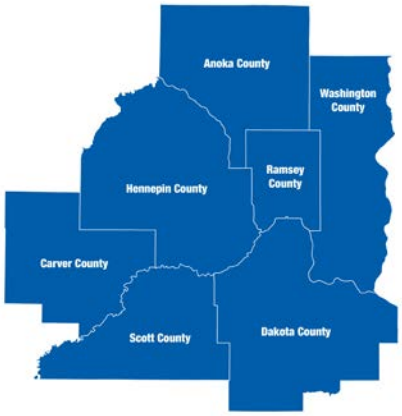
METROPOLITAN COUNCIL

April 2016

The Council's mission is to foster efficient and economic growth for a prosperous metropolitan region

Metropolitan Council Members

Adam Duininck	Chair	Edward Reynoso	District 9
Katie Rodriguez	District 1	Marie McCarthy	District 10
Lona Schreiber	District 2	Sandy Rummel	District 11
Jennifer Munt	District 3	Harry Melander	District 12
Deb Barber	District 4	Richard Kramer	District 13
Steve Elkins	District 5	Jon Commers	District 14
Gail Dorfman	District 6	Steven T. Chávez	District 15
Gary L. Cunningham	District 7	Wendy Wulff	District 16
Cara Letofsky	District 8		



The Metropolitan Council is the regional planning organization for the seven-county Twin Cities area. The Council operates the regional bus and rail system, collects and treats wastewater, coordinates regional water resources, plans and helps fund regional parks, and administers federal funds that provide housing opportunities for low- and moderate-income individuals and families. The 17-member Council board is appointed by and serves at the pleasure of the governor.

On request, this publication will be made available in alternative formats to people with disabilities. Call Metropolitan Council information at 651-602-1140 or TTY 651-291-0904.

This technical appendix describes the regression model in our *MetroStats*, "Diving Deeper: Understanding Disparities between Black and White Residents in the Twin Cities Region" (*PDF*). If you have additional questions, please contact us at research@metc.state.mn.us.

General notes

Unless otherwise noted, the source for all analyses in this report was the U.S. Census Bureau's American Community Survey (ACS) Public Use Microdata Sample (PUMS).

All regression models were based on the newly released 2010-2014 five-year ACS PUMS and employed the successive difference replicate weights that accompany the PUMS files. These weights reflect the complex sampling design of the ACS and result in more accurate statistical inference.

Age is modeled as a continuous variable with a quadratic term (age-squared). This reflects the rise in employment/income/homeownership as people reach middle age, and the decline in employment/income/homeownership after people retire.

We tested alternative specifications of certain control variables; results did not differ substantially:

- Using potential work experience (age minus years of education minus 6) instead of age. The patterns of racial disparities were very similar.
- Distinguishing people with graduate/professional degrees from people with a bachelor's degree alone. Differences between these two groups were not statistically significant.
- Distinguishing people with bachelor's degrees who majored in a field related to science, technology, engineering, or math from people who majored in another field. Differences between these two groups were not statistically significant.

All regression models are estimates that are subject to sampling error. Additionally, different tweaks to the models may yield slightly different results (though the overall story about disparities would not change). For these reasons, the numbers highlighted in the text should be interpreted as our best approximations of what is happening across the region, rather than exact or precise determinations.

To obtain the estimates of Black outcomes if Black demographics were the same as White demographics, we extracted Whites from the data and used the regression model coefficients to predict their outcomes as if they were Black (using SAS's PLM procedure). We then averaged the resulting predictions to arrive at the adjusted outcomes. For the employment and homeownership models, the predictions were in terms of probabilities that are readily interpretable. For the income models, the predictions were in terms of average logged wages; we exponentiated those averages to obtain the geometric means of hourly wages in dollars, which are more readily interpretable.

Notes about specific models appear below. Tables summarizing all models appear on the following pages.

Employment

We employed multinomial logistic regression models (SAS's SURVEYLOGISTIC procedure with a generalized logit link function) to analyze the likelihood of being: (a) employed; (b) unemployed; or (c) not in the labor force, the reference category. We restricted the analysis to Black and White non-institutionalized civilians between the ages of 16 and 64.

Income

Because the ACS asks only about total income in the past twelve months, and because not all respondents work all year, we converted annual wage/salary income to hourly earnings. To do this, we divided total wage/salary income in the past 12 months by the number of weeks the respondent worked in the past 12 months (to get the average weekly income), then divided by the usual number of hours worked in the past week.[1] The 2010-2014 ACS PUMS includes only ranges of weeks worked (e.g., 40-47 weeks); we used the continuous data on weeks worked from the 2005-2007 ACS PUMS to assign the average number of weeks worked within each range.

To mitigate any error introduced by this conversion, we winsorized the resulting distribution of hourly wages by bottom-coding at the 1st percentile and top-coding at the 99th percentile. That is, respondents whose hourly wages as calculated above were in the bottom 1 percent of hourly wages were assigned the 1st percentile of hourly wages, and respondents in the top 1 percent were assigned the 99th percentile. We also tested models of total wage/salary income, limiting the analysis to full-time, year-round workers; results were quite similar.

Before performing the regression (SAS's SURVEYREG procedure), we took the natural logarithm of hourly wages. This normalized the distribution, satisfying a key assumption of regression models.

We limited the analysis to Black and White non-institutionalized civilians between the ages of 25 and 64, with any wage/salary income in the past 12 months, who were employed in the private or public sector at the time they were surveyed. (This does not include self-employed people.)

Homeownership

We used standard logistic regression models (SAS's SURVEYLOGISTIC procedure with a cumulative logit link function) to examine the likelihood of owning one's home versus renting. An important control variable was household income. To obtain better estimates of its relationship to homeownership, and therefore obtain better estimates of the racial disparity in homeownership, we winsorized household income as described above for hourly wages. This resulted in considerably better model fit. We also took the natural logarithm of the resulting variable; this reflects how the effect of income on homeownership declines at higher income levels. That is, an extra $30,000 of income makes a big difference to homeownership if it takes

[1] Paid leave counts as work time.

a household from $30,000 to $60,000, but probably not a big difference if it takes a household from $170,000 to $200,000 (both of these higher-income households have enough money to afford a mortgage).

We restricted the analysis to Black and White household

Regression Models of Employment Status (comparison: employed versus not in labor force)

	Model #1			Model #2			Model #3	
	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.
Intercept	1.587***	(0.015)		1.192***	(0.113)		-3.905***	(0.122)
Black, non-Latino	-0.667***	(0.049)		-0.795***	(0.062)		-0.339***	(0.064)
Age				0.349***	(0.007)		0.286***	(0.007)
Age-squared				-0.004***	(8.6e-5)		-0.004***	(9.2e-5)
Speak English less than "very well"				-0.477***	(0.116)		-0.032	(0.135)
Nativity/length of residence (reference = native-born)								
Born outside US, < 5 years in US				-0.606***	(0.154)		-0.875***	(0.160)
Born outside US, 5-9 years in US				0.286***	(0.197)		0.027	(0.217)
Born outside US, 10-14 years in US				0.303***	(0.159)		0.067	(0.171)
Born outside US, 15+ years in US				0.453***	(0.122)		0.208	(0.133)
Female							-0.508***	(0.034)
Lived outside the region one year ago							-0.608***	(0.082)
Has a disability							-0.178***	(0.049)
Has child under age 6							-0.524***	(0.053)
Highest degree earned (reference = no high school diploma)								
High school diploma or equivalent							1.255***	(0.062)
Associate's degree							1.706***	(0.077)
Bachelor's degree or higher							1.702***	(0.062)

Note: Entries are coefficients from multinomial logistic regression models. Replicate standard errors are in parentheses.

*** $p < 0.001$ (two-tailed tests).

Regression Models of Employment Status (comparison: unemployed versus not in labor force)

	Model #1			Model #2			Model #3	
	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.
Intercept	-1.270***	(0.030)		-4.431***	(0.205)		-4.125***	(0.210)
Black, non-Latino	0.643***	(0.070)		0.585***	(0.084)		0.757***	(0.088)
Age				0.190***	(0.012)		0.199***	(0.015)
Age-squared				-0.003***	(1.5e-4)		-0.003***	(1.8e-4)
Speak English less than "very well"				-0.157	(0.149)		0.057	(0.163)
Nativity/length of residence (reference = native-born)								
Born outside US, < 5 years in US				-0.555*	(0.223)		-0.793**	(0.242)
Born outside US, 5-9 years in US				-0.076	(0.239)		-0.188	(0.240)
Born outside US, 10-14 years in US				0.046	(0.237)		-0.006	(0.246)
Born outside US, 15+ years in US				-0.233	(0.177)		-0.325	(0.186)
Female							-0.623***	(0.056)
Lived outside the region one year ago							0.068	(0.158)
Has a disability							-1.091***	(0.090)
Has child under age 6							-0.705***	(0.104)
Highest degree earned (reference = no high school diploma)								
High school diploma or equivalent							0.666***	(0.102)
Associate's degree							0.691***	(0.151)
Bachelor's degree or higher							0.256*	(0.103)

Note: Coefficients are from multinomial logistic regression models. Standard errors are based on the Census Bureau's successive difference replicate weights.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed tests).

Regression Models of Logged Hourly Wages

	Model #1			Model #2			Model #3	
	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.
Intercept	3.169***	(0.005)		1.192***	(0.060)		1.289***	(0.171)
Black, non-Latino	-0.402***	(0.021)		-0.380***	(0.024)		-0.148***	(0.021)
Age				0.085***	(0.003)		0.066***	(0.003)
Age-squared				-8.6e-4***	(<1e-4)		-6.0e-4***	(<1e-4)
Speak English less than "very well"				-0.259***	(0.043)		-0.047	(0.040)
Nativity/length of residence (reference = native-born)								
Born outside US, < 5 years in US				0.029	(0.059)		0.023	(0.056)
Born outside US, 5-9 years in US				0.060	(0.045)		-0.018	(0.046)
Born outside US, 10-14 years in US				0.121**	(0.039)		0.012	(0.039)
Born outside US, 15+ years in US				0.131***	(0.032)		0.002	(0.026)
Female							-0.161***	(0.007)
Lived outside the region one year ago							-0.092**	(0.027)
Has a disability							-0.192***	(0.020)
Has child under age 6							0.137***	(0.012)
Highest degree earned (reference = no high school diploma)								
High school diploma or equivalent							0.136***	(0.033)
Associate's degree							0.210***	(0.034)
Bachelor's degree or higher							0.453***	(0.035)
Employed full-time, year-round							0.211***	(0.010)
Class of worker (reference = private for-profit)								
Private non-profit							-0.080***	(0.012)
Public							0.003	(0.011)

Note: Standard errors are based on the Census Bureau's successive difference replicate weights. Fixed effects for occupations (23 two-digit Standard Occupational Classification codes) are also included but not shown.

** $p < 0.01$; *** $p < 0.001$ (two-tailed tests).

How to interpret this table:

- Positive numbers indicate that the given characteristic has a positive effect on wages after accounting for the influence of the other factors in the model; negative numbers indicate that the given characteristic has a negative effect on wages after controlling for the other factors. For example: the negative sign for "Black, non-Latino" in Model #3 indicates that relative to Whites, Blacks have lower wages, controlling for all other factors listed above.

- Asterisks show how confident we can be that the given characteristic has an effect in the population, after controlling for the other characteristics. The more asterisks by a given number, the more confident we can be. For example: the three asterisks by “Black, non-Latino” in Model #3 indicate that we can be extremely (99.9%) confident that Blacks’ average wages differ from Whites’ average wages in the region, even after accounting for the influence of the other factors in the model.
- Some characteristics have more than two categories. In these cases, effects are stated relative to the “reference” category. For example: the positive sign of the coefficient for “High school diploma or equivalent” in Model #3 indicates that wages for workers with a high school diploma are higher than for workers without a high school diploma, even after accounting for the influence of other factors.

Regression Models of Homeownership (comparison: owning versus renting)

	Model #1			Model #2			Model #3	
	Coeff.	S.E.		Coeff.	S.E.		Coeff.	S.E.
Intercept	1.133***	(0.015)		-5.199***	(0.140)		-14.660***	(0.369)
Black, non-Latino	-2.239***	(0.059)		-2.214***	(0.074)		-1.679***	(0.076)
Age				0.240***	(0.006)		0.207***	(0.006)
Age-squared				-0.002***	(5.0e-5)		-0.001***	(5.5e-5)
Speak English less than "very well"				-0.897***	(0.166)		-0.334*	(0.162)
Nativity/length of residence (reference = native-born)								
Born outside US, < 5 years in US				-1.203**	(0.381)		-1.151**	(0.442)
Born outside US, 5-9 years in US				-0.422	(0.254)		-0.736**	(0.259)
Born outside US, 10-14 years in US				0.118	(0.168)		-0.219	(0.173)
Born outside US, 15+ years in US				0.478***	(0.118)		0.226	(0.122)
Female							-0.178***	(0.037)
Lived outside the region one year ago							-1.846***	(0.123)
Has a disability							-0.777***	(0.059)
Has child under age 6							0.789***	(0.069)
Highest degree earned (reference = no high school diploma)								
High school diploma or equivalent							0.362***	(0.093)
Associate's degree							0.667***	(0.100)
Bachelor's degree or higher							0.673***	(0.094)
Number of full-time, year-round workers in household (reference = 0)								
1							-0.046	(0.052)
2 or more							0.377***	(0.080)
Household income (ln)							0.846***	(0.034)

Note: Coefficients are from logistic regression models. Standard errors are based on the Census Bureau's successive difference replicate weights.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (two-tailed tests).

390 Robert Street North
Saint Paul, MN 55101-1805

651.602.1000
TTY 651.291.0904
research@metc.state.mn.us
metrocouncil.org/data

Follow us on:
twitter.com/metcouncilnews
facebook.com/MetropolitanCouncil
youtube.com/MetropolitanCouncil