PUBLIC POLICY INSTITUTE OF CALIFORNIA

California's Aging Population

Anticipating Dramatic Growth in the Number of Older Californians

Technical Appendices

CONTENTS

Hans Johnson, Eric McGhee, Paulette Cha, and Shannon McConville

with research support from Shalini Mustala

Appendix A. Predictive data and methods

Our methodological approach uses Department of Finance population projections as our base and attaches characteristics of interest to those projections.

Every few years, or so, the California Department of Finance produces long-term population projections for the state. The most recent series is considered an "interim" set of projections, and will be replaced soon ("summer 2024") with a new completely revised projection that relies on the 2020 Census as the base year. The interim projections include race/ethnicity, single year of age, and gender, and extend to 2060. The Department of Finance series are the official state projections to be used by state agencies in their long-term planning processes.

Because the Department of Finance projections do not include key characteristics of the older population that are useful for policy and practice, we develop our own set of projections of those characteristics. Specifically, we project estimated probabilities that populations by age, race/ethnicity, and gender will have a particular characteristic of interest, such as college attainment or homeownership. We use machine learning models to generate projections based on trends identified in the American Community Surveys from 2006 through 2022.

If two data sets emerge from the same fundamental data generating process, errors in predicting from one data set to the other come from three separate sources. *Bias* is the error from reducing the complexity of the world to a relatively simple model. *Variance* is the degree to which a model's results differ when fit to different data sets. The remaining error is the unavoidable randomness of any data source.

Bias and variance are often at odds with each other. Simple models—the mean of the data set, for example, or a straight line—are relatively insensitive to randomness in the data and will produce consistent predictions from one data set to the next. They also may suffer from high bias because they miss important nuances in the relationships being modeled. Yet as models become more complex to capture these nuances, they capture more one-off deviations that have nothing to do with those same fundamental relationships. At the extreme, a model that perfectly predicts every outcome in one data set will miss any difference (no matter how slight) between that data set and a different one created by the same data generating process. The goal is to increase complexity to reduce bias, but only to the point that it does not become dwarfed by variance.

Our prediction method is Gradient Boosted Decision Trees. Gradient Boosted Decision Tree (GBDT) models are a powerful and increasingly popular machine learning technique. GBDT models combine the strengths of decision trees with a method called gradient boosting to create predictive models.

Decision trees are intuitive models that make predictions by asking a series of yes/no questions about the input data, similar to a flowchart. The user identifies an outcome variable to be predicted and a set of input variables that will serve as the basis for the prediction. The algorithm starts by splitting the data on the single input variable and its partition that minimize a specified loss function (i.e., a measure of predictive error) for each resulting subset of the data. Our outcome variables are all binary, so we use the log loss, $-[y\, log(p) + (1-y)\, log(1-p)]$, where y is the actual outcome (coded 1 for individuals with the characteristic and 0 otherwise) and p is the predicted probability for that outcome. The process repeats for each subset, identifying the optimal input variable and its split (sometimes another split of the same one) until some pre-determined level of complexity is reached.

While individual decision trees are easy to understand, they suffer from drawbacks. First, in order to make the problem tractable, the decision tree algorithm is "greedy": it takes the best split at each step in the process, without exploring if that split leads to an inferior split later on. Second, individual trees can suffer from high variance because they are optimized for the data at hand and so represent just one possible model of those data.

Gradient boosting addresses these limitations with an ensemble of many trees. It does this iteratively, with each new tree focusing on predicting the mistakes—the residuals—of the previous trees. This process can be represented by the following equation:

$$F(x) = \sum_{t=1}^{T} \lambda h_t(x)$$

Where:

- $F(x)$ is the final prediction
- T is the number of trees (or rounds of model updates)
- λ is the learning rate (a small number that controls how quickly the model learns)
- $h_t(x)$ is the prediction of the t-th tree.

The model starts with a simple prediction (often just the average of the outcome variable) and then adds trees one by one. Each new tree is trained to predict the residuals (the differences between the current predictions and the actual values) of the current model. These predictions are then added to the model, scaled by the learning rate λ.

This iterative process continues until a specified number of trees is reached or until the model's performance stops improving. The learning rate λ is typically set to a small value (e.g., 0.1 or smaller) to prevent overfitting, which occurs when a model performs well on training data but poorly on new, unseen data.

There are three main parameters in GBDT that control how complex the model is allowed to become. In addition to T, the number of total trees, and λ, the rate at which the predictions are updated, there is d, the number of splits permitted for each tree. The higher the value of d, the more complex each tree.

Complexity is only an asset if the true relationship we are modeling is also complex. To find the optimal balance between bias and variance, we used the following process:

1. Cycle through 60 combinations of T, λ, and d.
    a. $T \in \{10, 20, 30, 40, 50\}$
    b. $\lambda \in \{0.001, 0.01\}$
    c. $d \in \{1, 2, 3, 4, 5, 6\}$
2. For each combination, run k-fold cross-validation, with $k = 10$.
    a. Split the data into 10 random subsets.
    b. Use each subset to test the model and train the model on the remaining 90% of the data. Predictors include age (single year), birth year, race (dummy variables for non-Hispanic white, Black, Asian American, Latino, and Pacific Islander), and gender.
    c. Calculate the log loss of the predictions on that test data set.
    d. Average the log loss across all 10 subsets.
3. Identify the combination with the best average log loss and use that for all predictions.

The data for modeling come from the public use microdata files of the American Community Survey (ACS) of the U.S. Census, as provided by IPUMS at the University of Minnesota (https://usa.ipums.org/usa/). The ACS is a large-scale survey of about 3 million American households conducted on a rolling basis throughout each year and aggregated to a single data set for public release. The PUMS data are compilations of characteristics with weights to match the corresponding number of Americans in the general population. We draw a weighted sample of one million cases from the PUMS data (about one fifth of the total household sample for California) and model with that data set.

Because GBDT modeling is not the only approach to prediction, Table B1 compares the overall log loss for each of our outcomes to the log loss from two alternative prediction methods: the sample mean, and a logistic regression with the same predictors as the GBDT model. As before, the model is calculated using 80% of the data for training and predicted with the remaining 20%. In all cases, our chosen GBDT model has the lowest out-of-sample log loss. The advantage over the sample mean is often substantial, but the benefit over a logistic regression is generally far smaller. We also explored using five-year age groups for our modeling to ensure more stability in the survey estimates, but it produced no improvement in out of sample predictive accuracy (results available from authors upon request).

The figures in Appendix B plot the predicted and actual proportions by age for each outcome variable, separately by year, race, and gender. In Figures B1-B42, we trained the model from the k-fold cross-validation process described above on a random subset of 80% of the data, and then tested it on the remaining 20%. Because our ultimate goal was an analysis of ages 55 and over and because we included birth year as an explanatory variable, we modeled using ages 37 and over so that everyone in our sample in 2022 would be at least 55 in 2040. However, this validation process arguably overstates confidence in predicting time-dependent variation in the data, since the test data control for any unmeasured sources of variation over time. Thus, in Figures B43-B84 we trained the model on all of the data from 2006 through 2014 (i.e., not a random subset), and then tested on 2015 through 2022. As we discuss in the main text, the predictive accuracy was lower for certain key outcomes that are likely to vary over time, such as income and homeownership.

Appendix B. Predictive validation

TABLE B1

Log loss of alternative predictive models compared to gradient boosted decision trees

Outcome	GBDT	Sample mean	Logistic regression
Less than high school graduate	0.393	0.487	0.396
High school graduate	0.502	0.509	0.506
Some college	0.582	0.597	0.585
College graduate	0.563	0.627	0.567
Married	0.650	0.682	0.664
U.S. born	0.451	0.663	0.454
Foreign language at home	0.412	0.680	0.414
Homeowner	0.611	0.645	0.619
Income < 2 X poverty threshold	0.531	0.559	0.536
In the labor force	0.486	0.673	0.507
Live alone	0.373	0.407	0.377
In institutional group quarters	0.095	0.104	0.099
Difficulty with independent living	0.230	0.275	0.233
Difficulty with self care	0.162	0.191	0.164

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTES: Cell entries show the overall log loss when predicting each outcome variable in the rows using the model identified in each column header. Training is done with a random 80% sample of the data, and the predictions are compared to the remaining 20%. Larger values indicate poorer model fit.

FIGURE B1

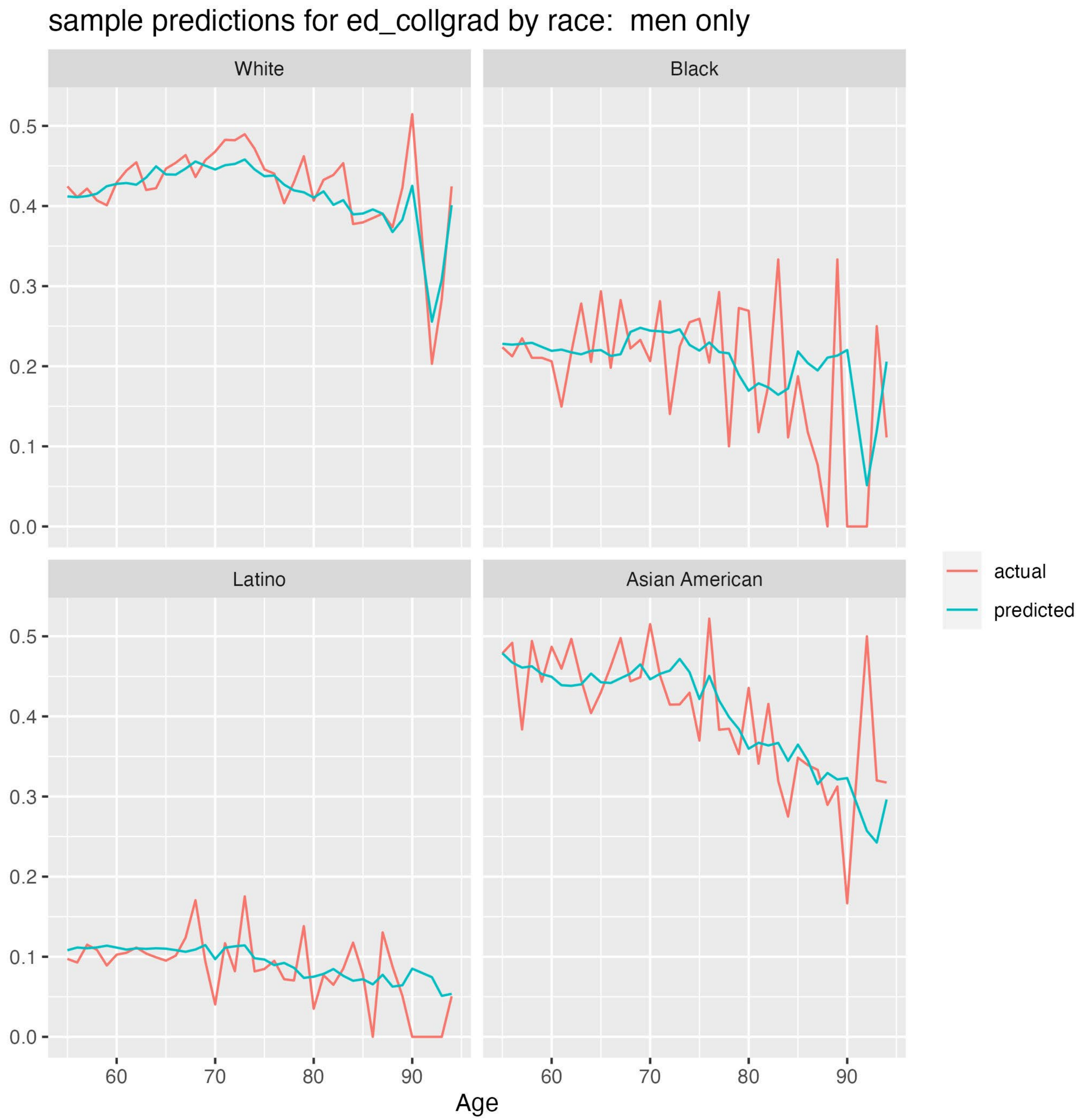
Predicted vs. actual from 80% training sample: college graduates by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B2

Predicted vs. actual from 80% training sample: college graduates by race, men only

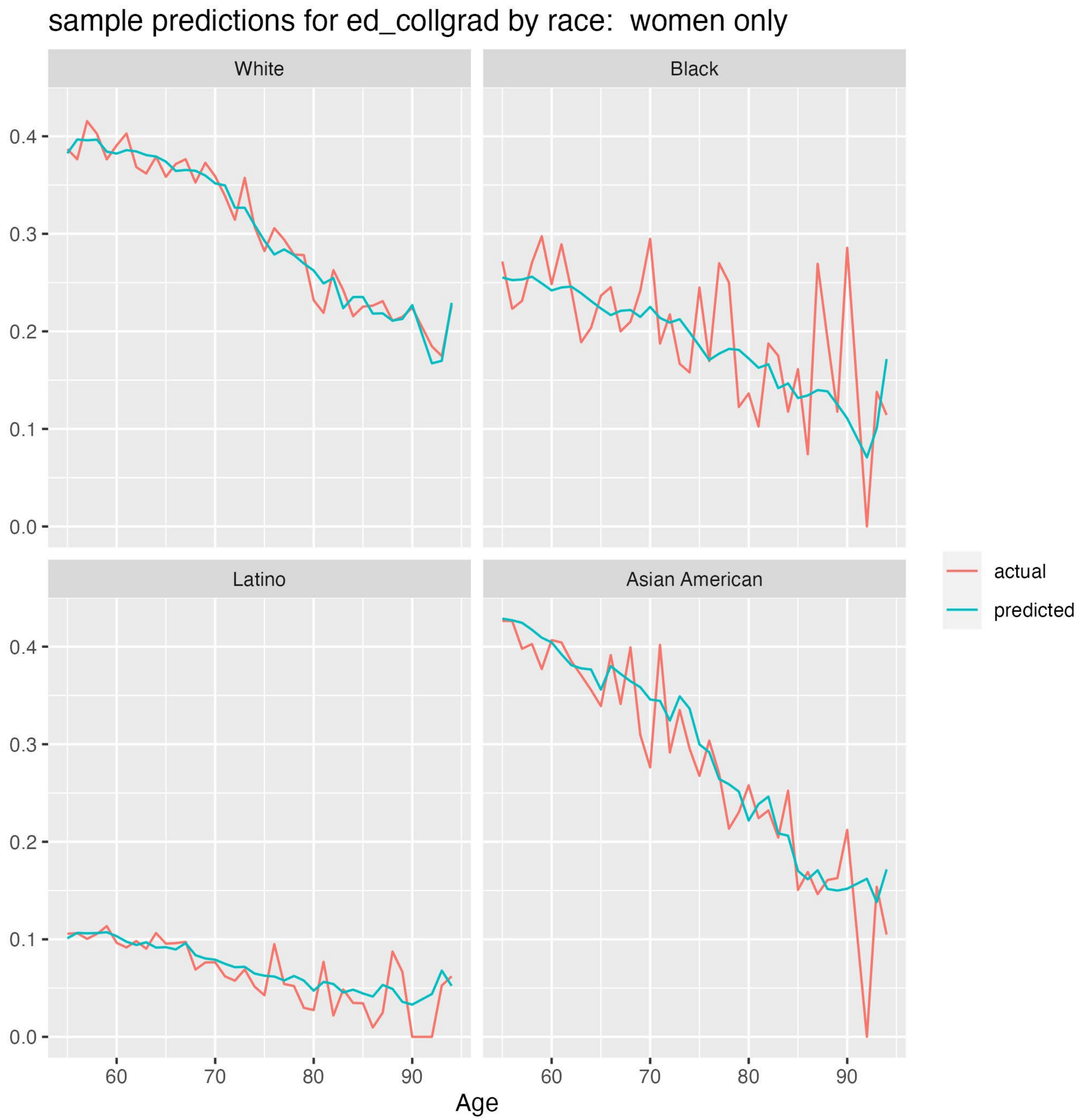


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B3

Predicted vs. actual from 80% training sample: college graduates by race, women only

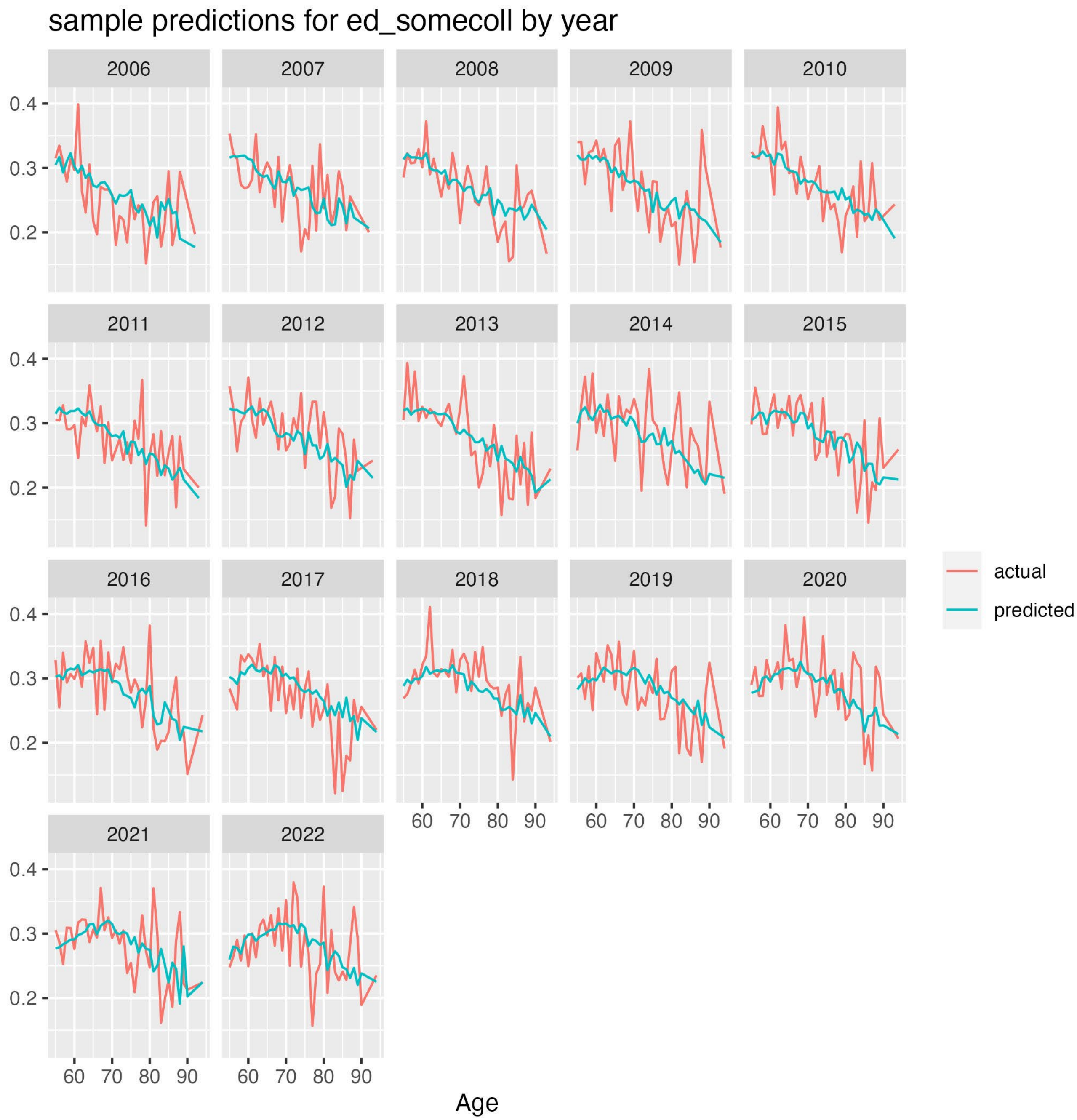


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B4

Predicted vs. actual from 80% training sample: some college by year

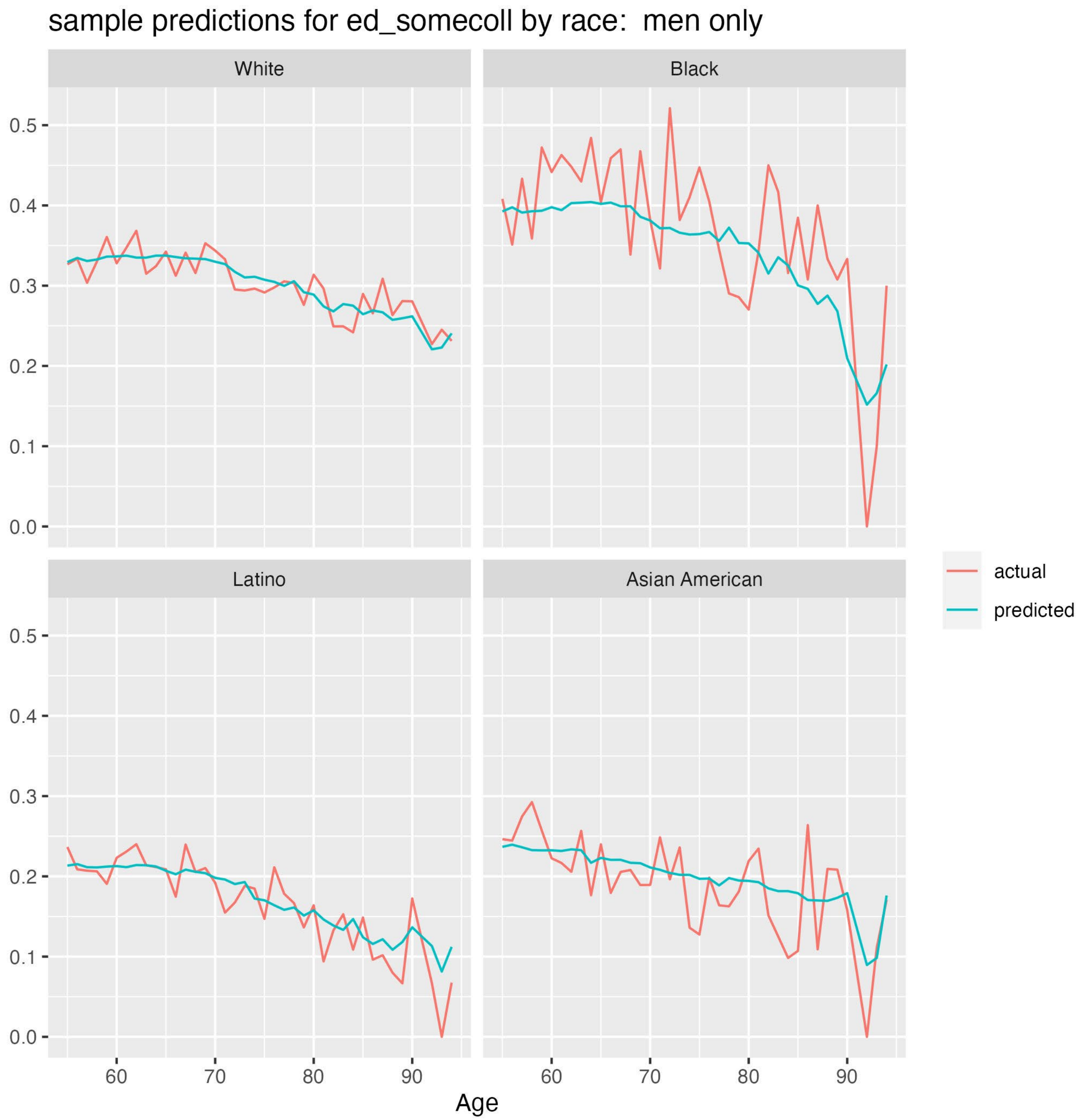


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B5

Predicted vs. actual from 80% training sample: some college by race, men only

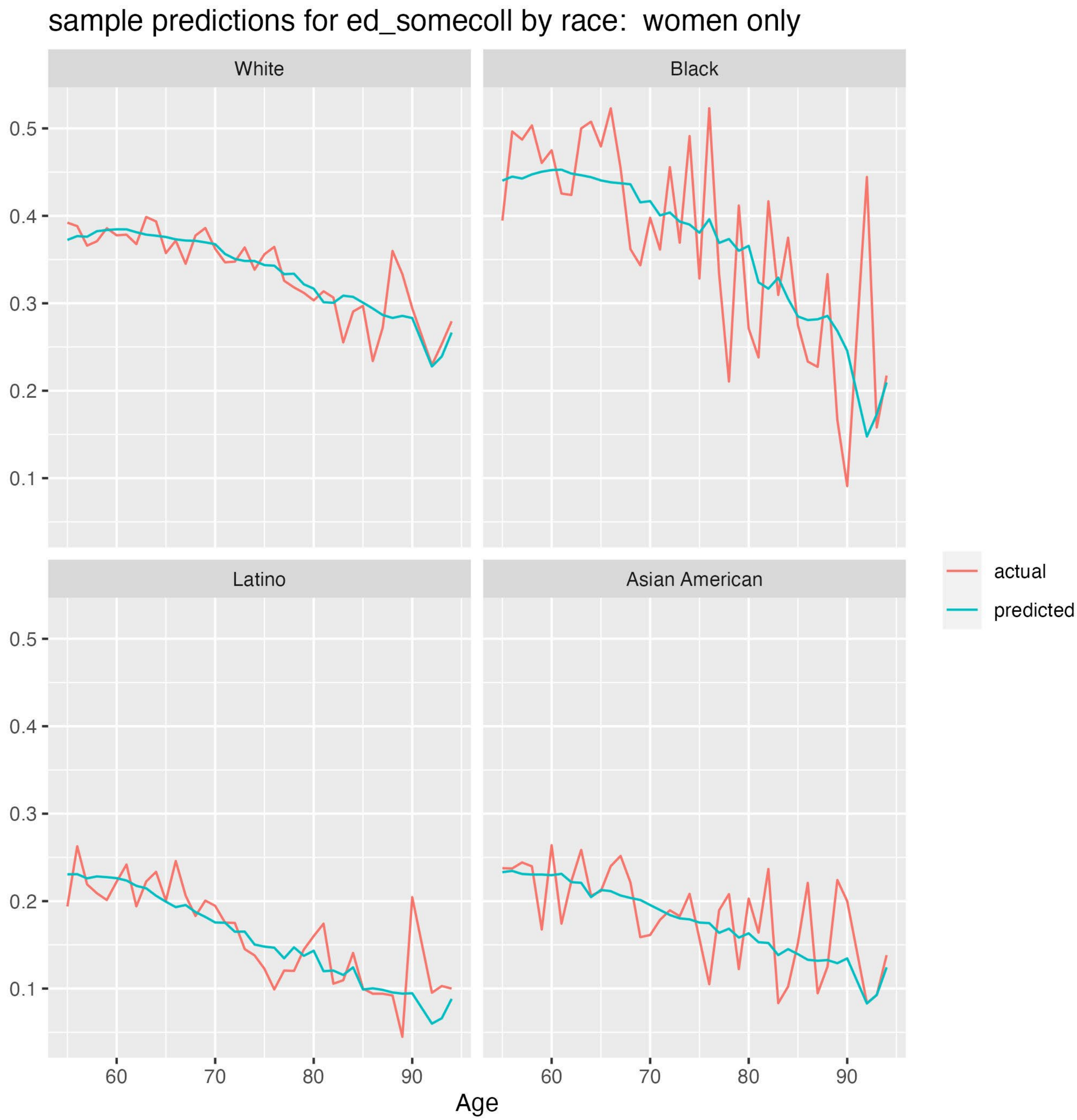


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B6

Predicted vs. actual from 80% training sample: some college by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B7

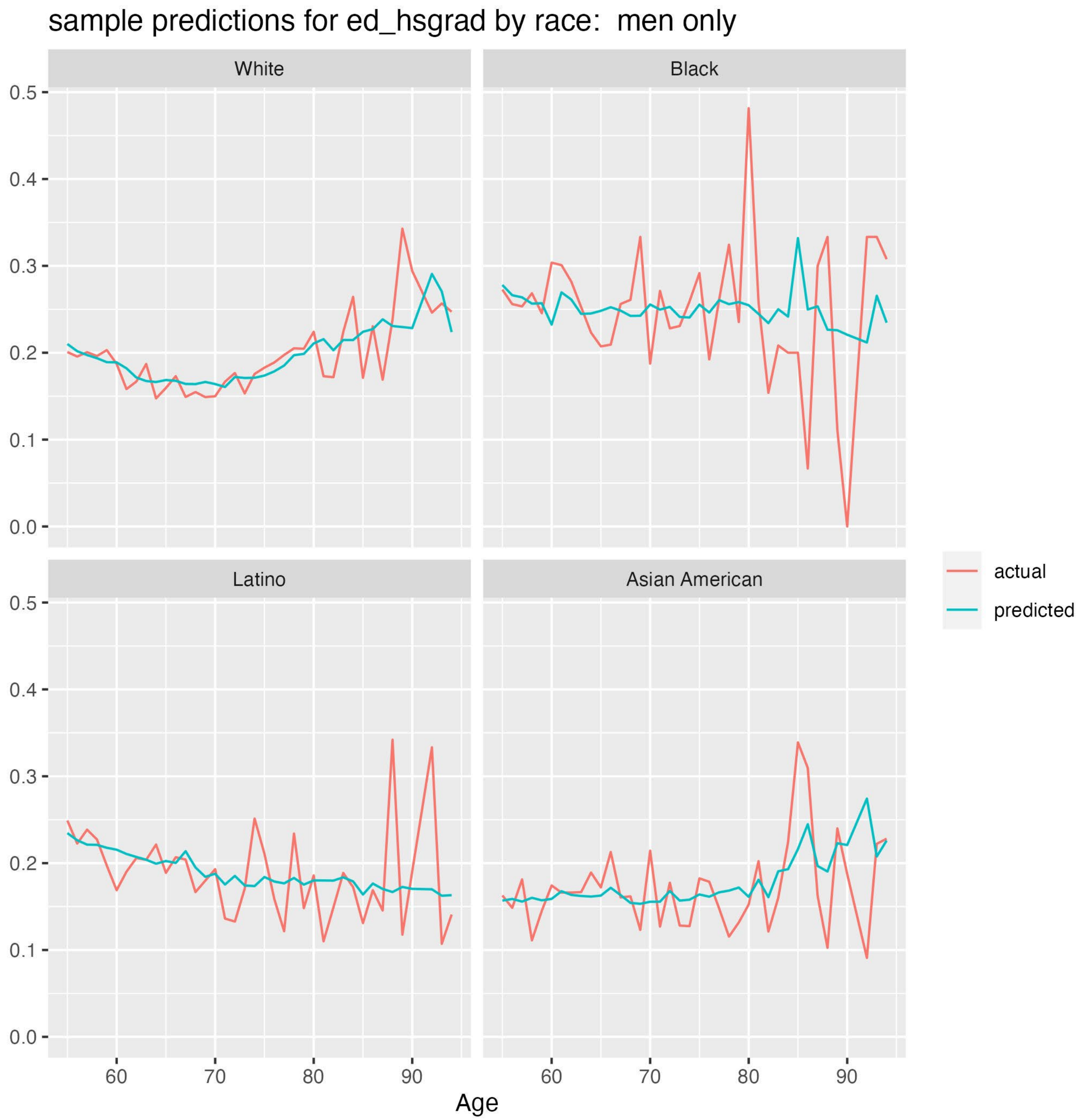
Predicted vs. actual from 80% training sample: high school graduates by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B8

Predicted vs. actual from 80% training sample: high school graduates by race, men only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B9

Predicted vs. actual from 80% training sample: high school graduates by race, women only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B10

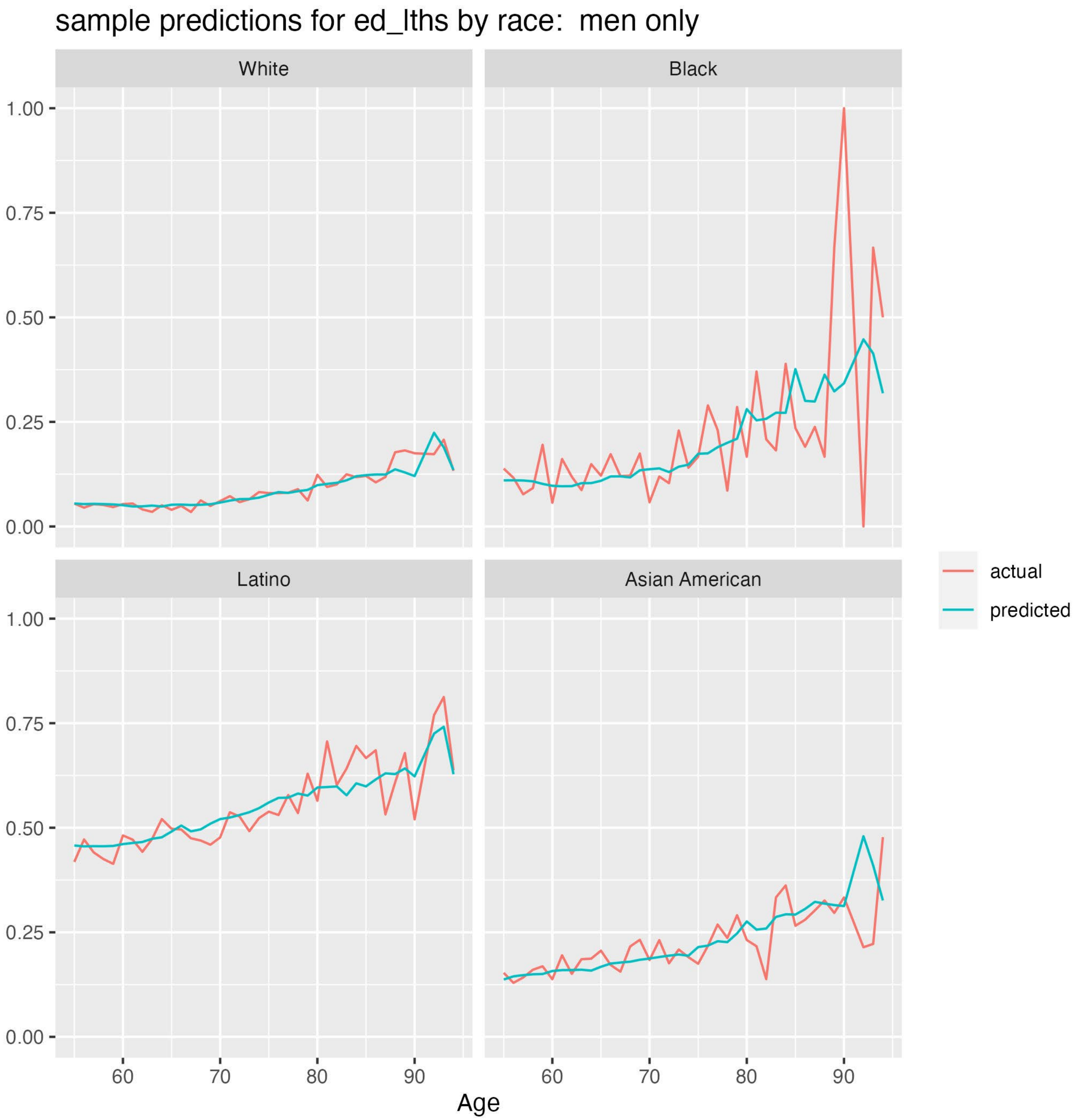
Predicted vs. actual from 80% training sample: less than high school by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B11

Predicted vs. actual from 80% training sample: less than high school by race, men only

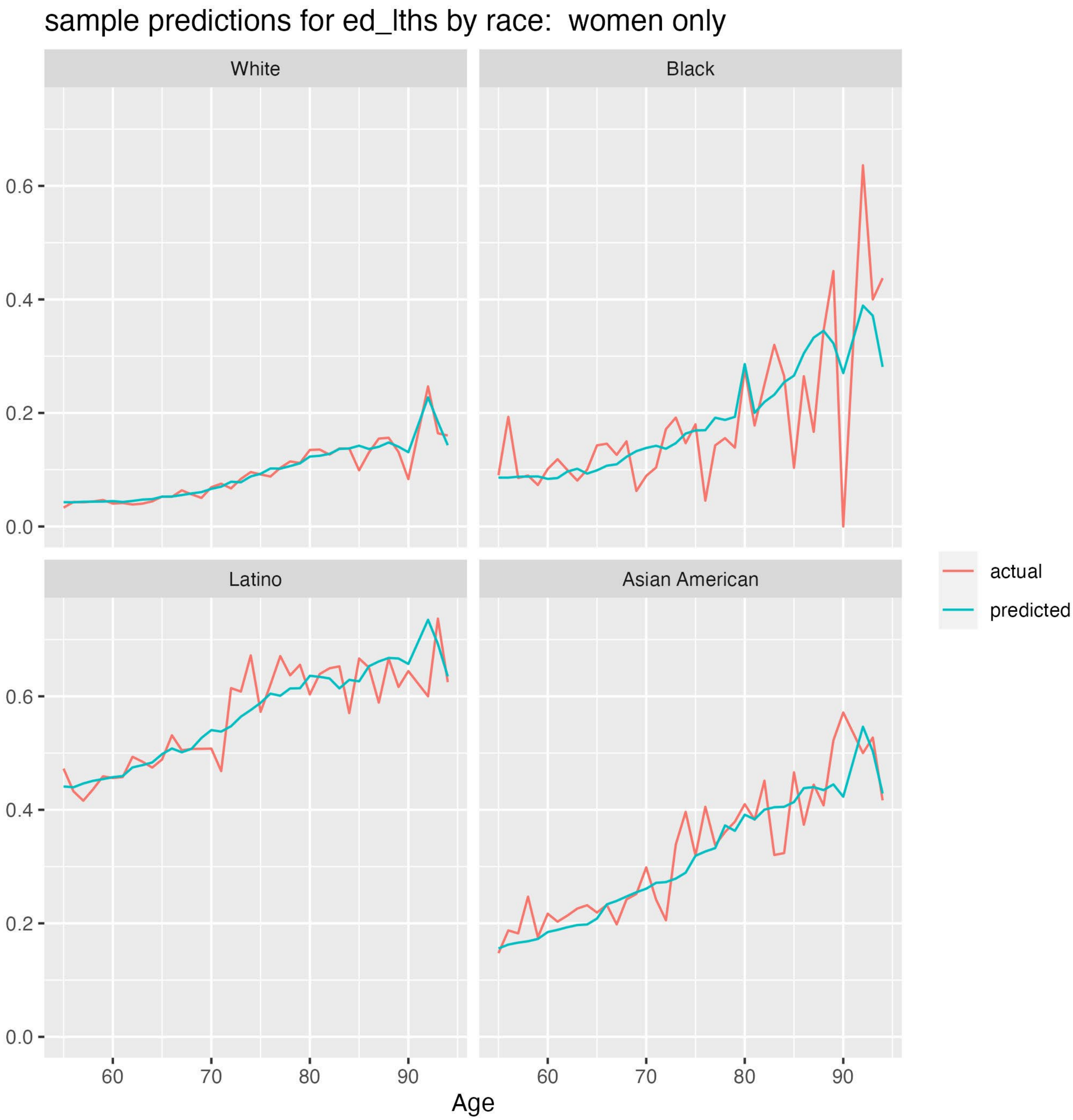


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B12

Predicted vs. actual from 80% training sample: less than high school by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B13

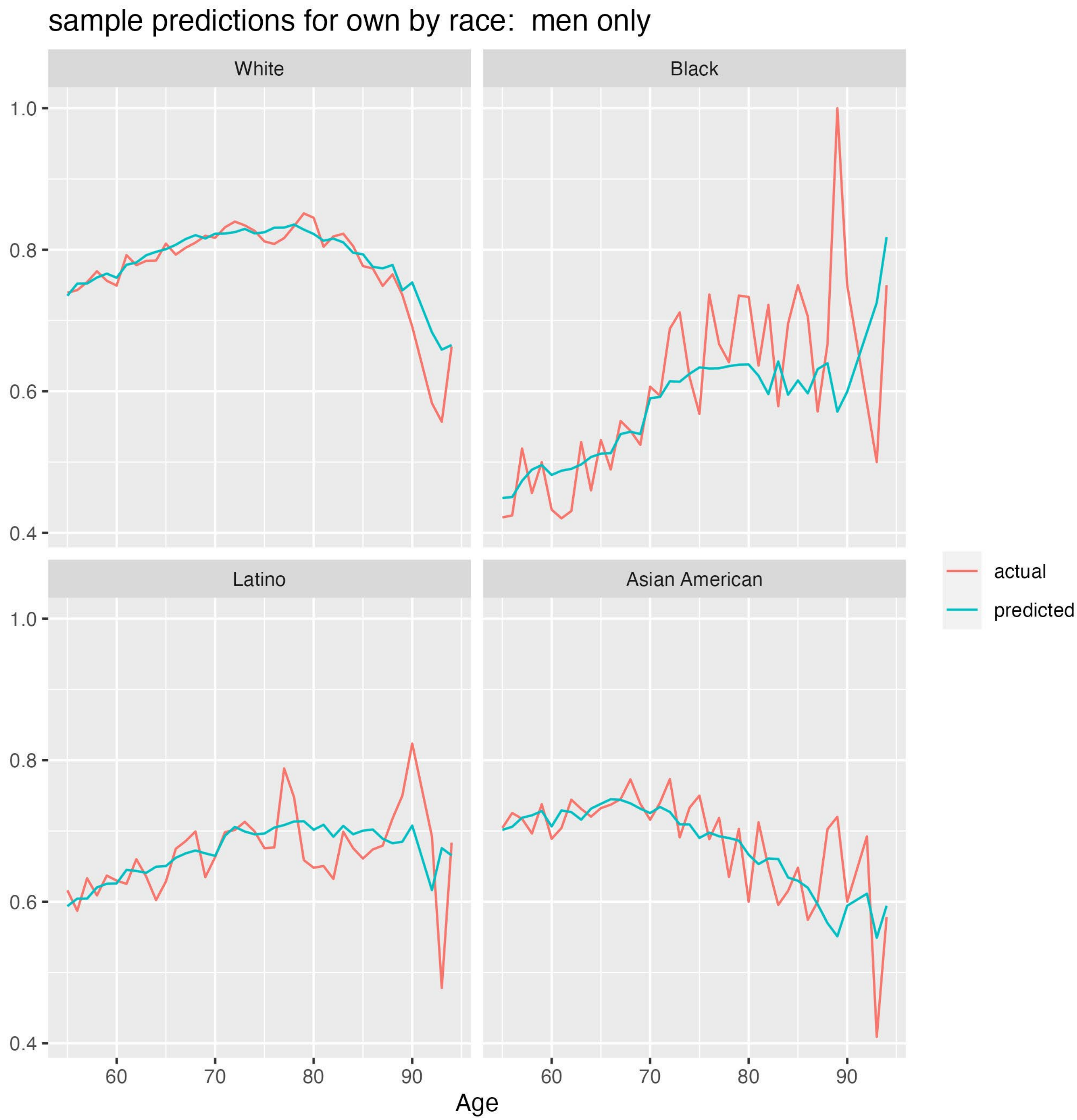
Predicted vs. actual from 80% training sample: home ownership by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B14

Predicted vs. actual from 80% training sample: home ownership by race, men only

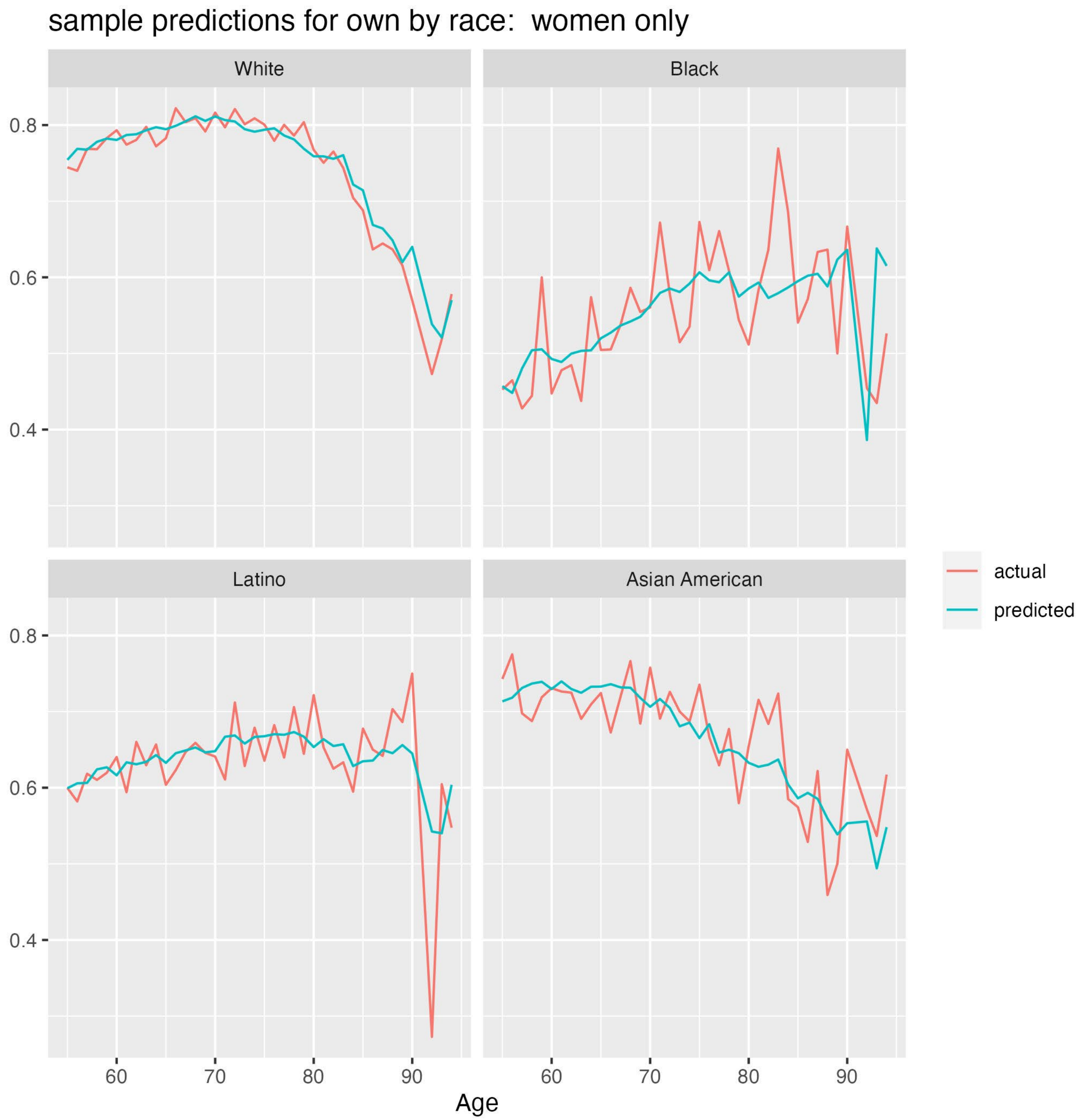


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B15

Predicted vs. actual from 80% training sample: home ownership by race, women only

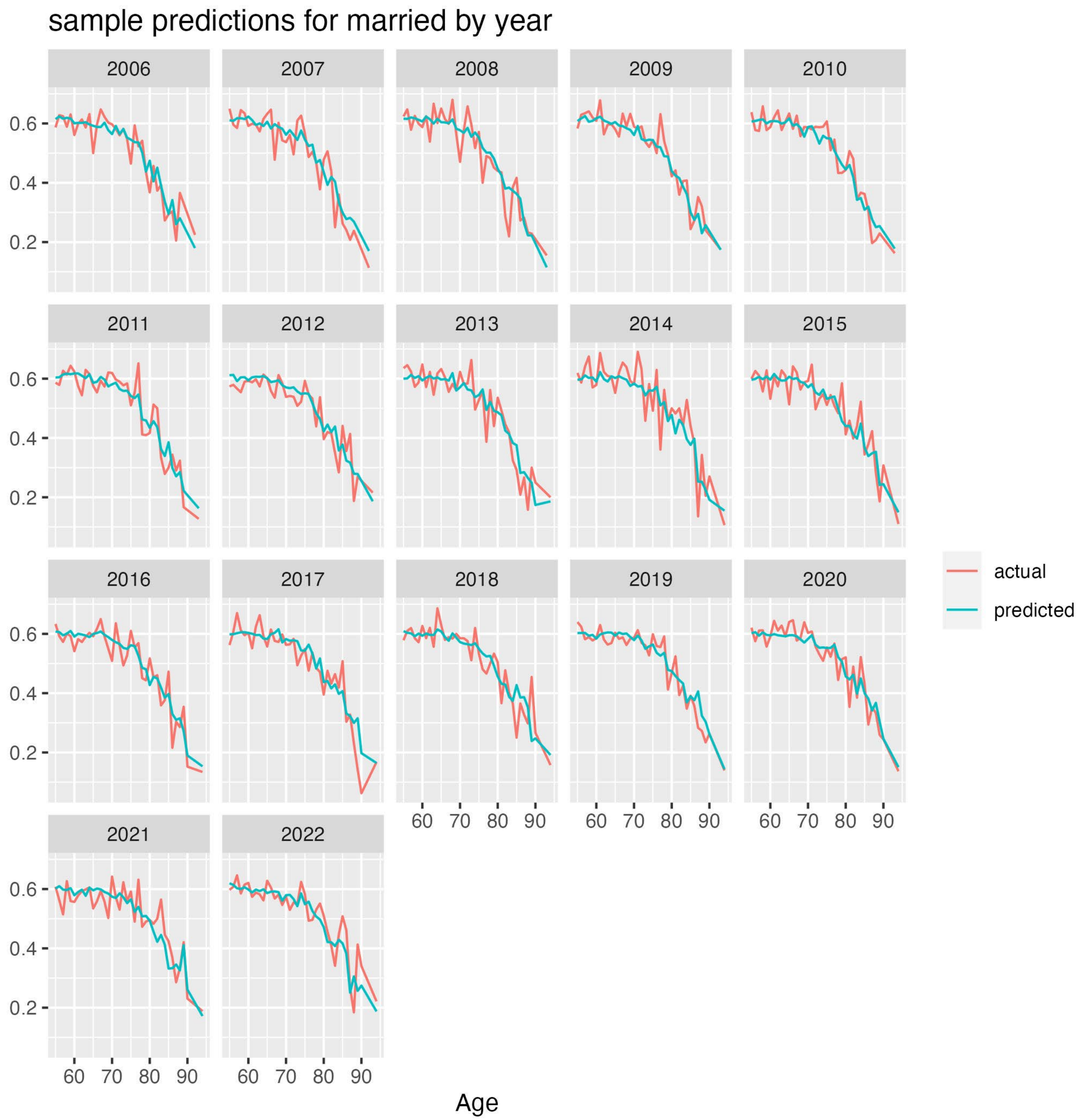


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B16

Predicted vs. actual from 80% training sample: married by year

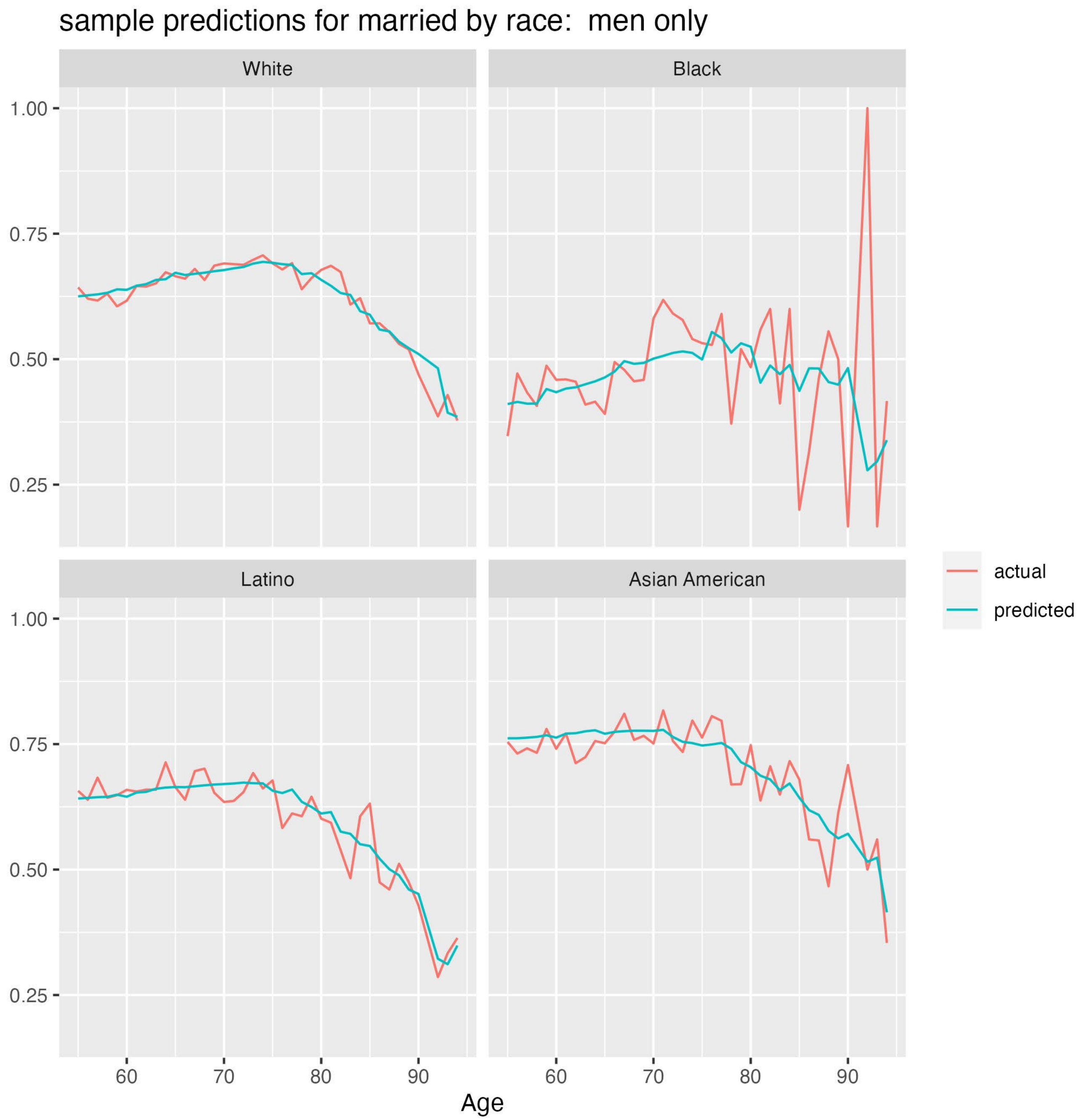


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B17

Predicted vs. actual from 80% training sample: married by race, men only

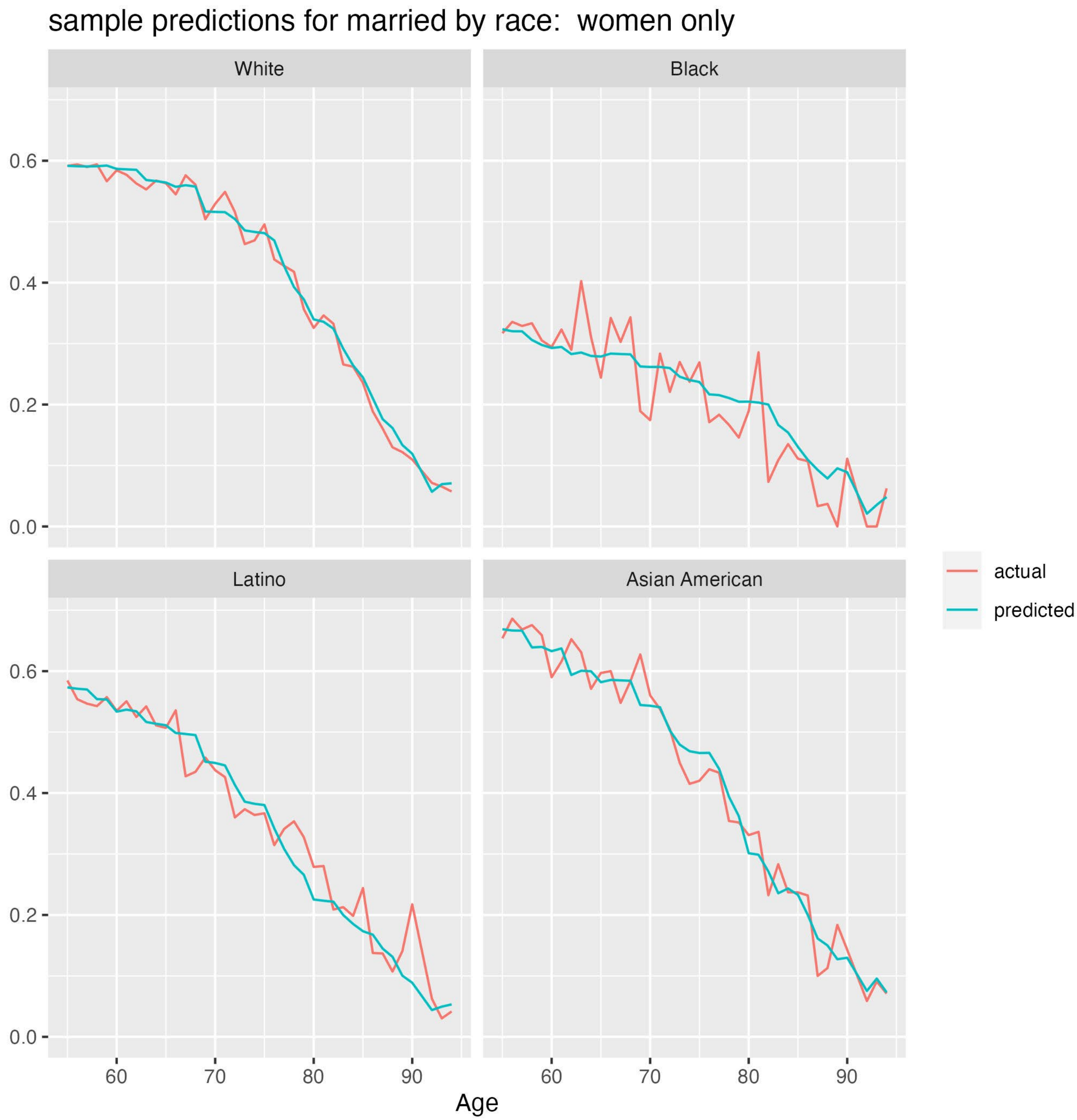


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B18

Predicted vs. actual from 80% training sample: married by race, women only

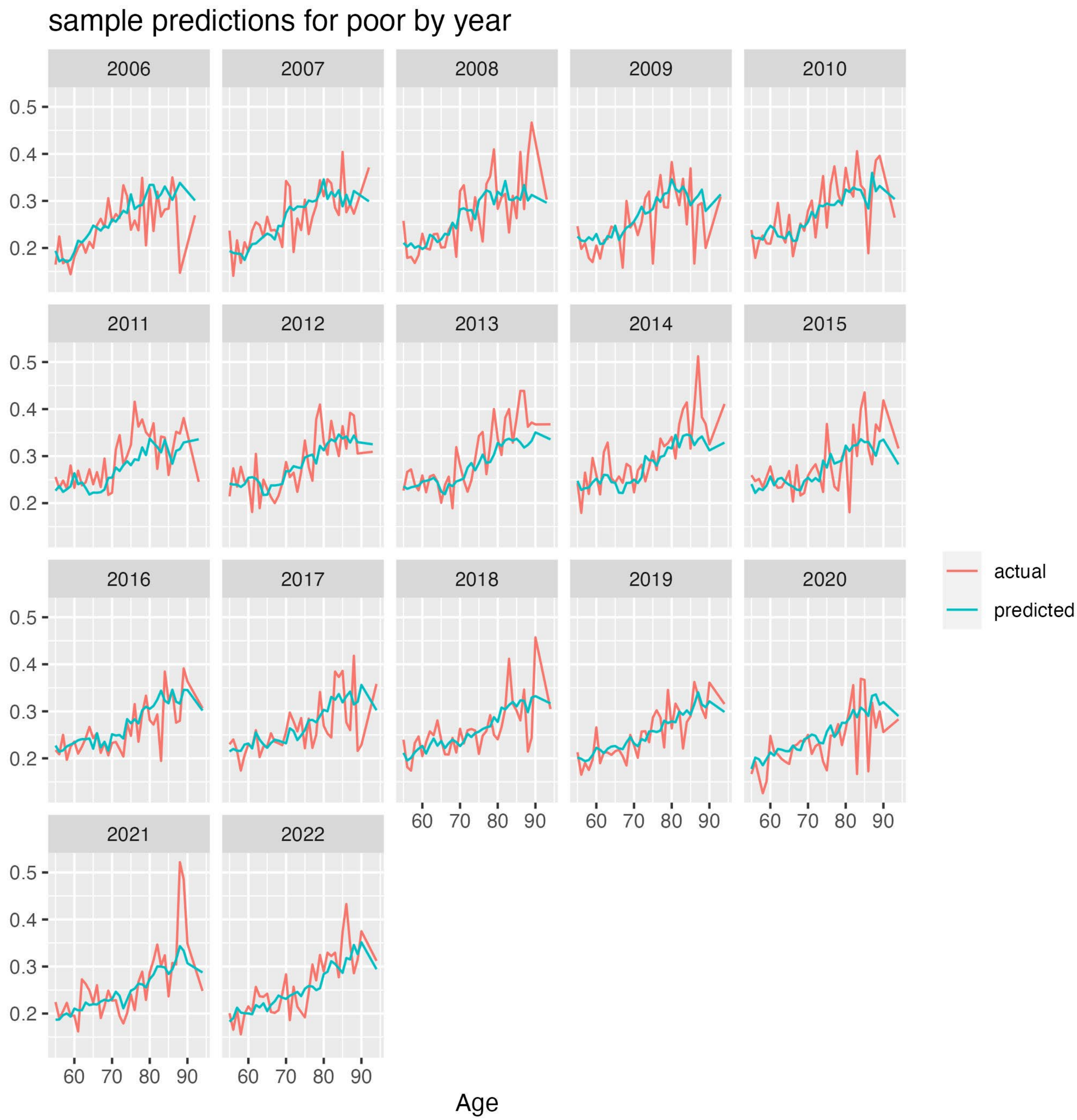


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B19

Predicted vs. actual from 80% training sample: poor (less than twice poverty level) by year

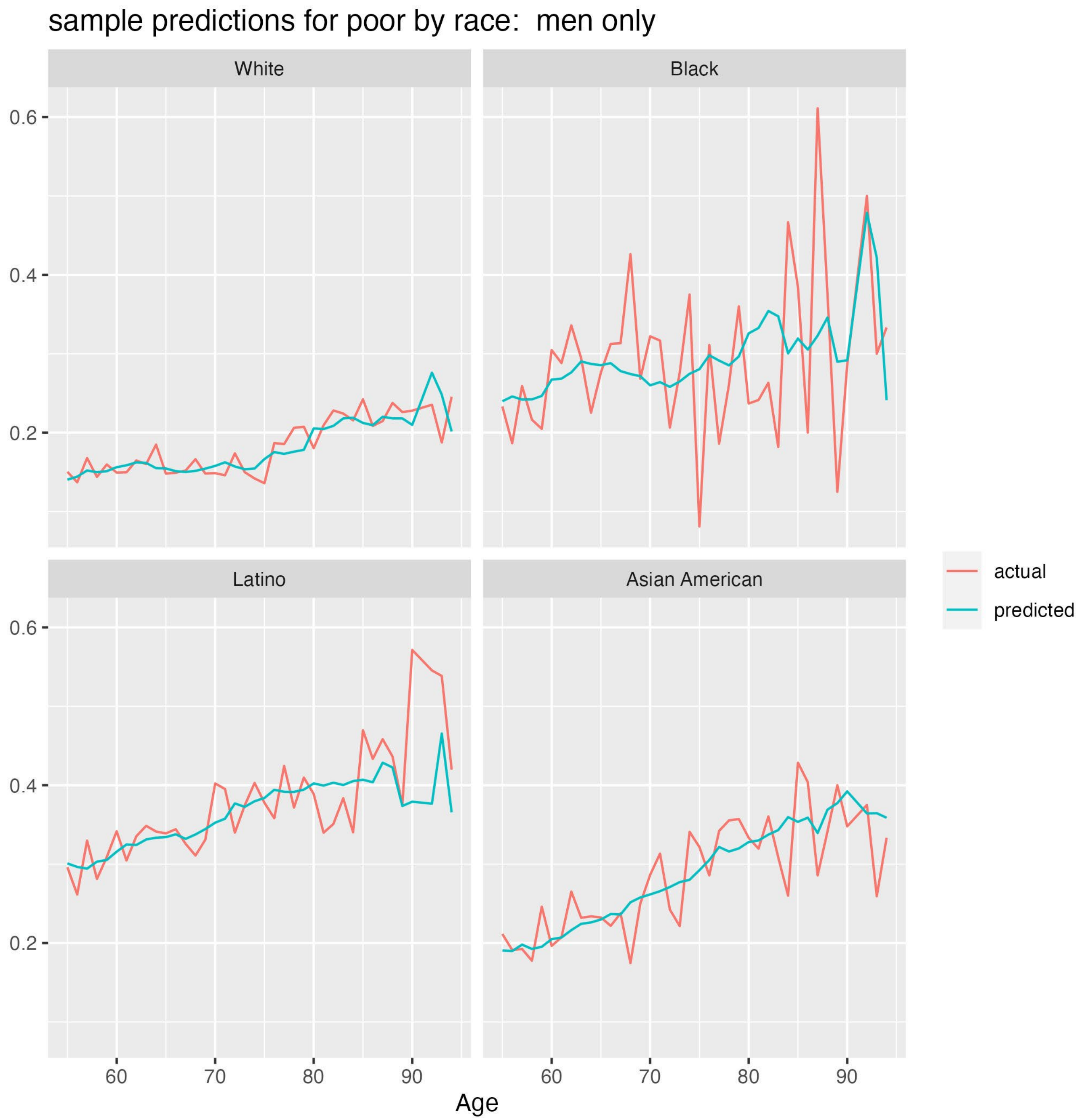


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B20

Predicted vs. actual from 80% training sample: poor (less than twice poverty level) by race, men only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B21

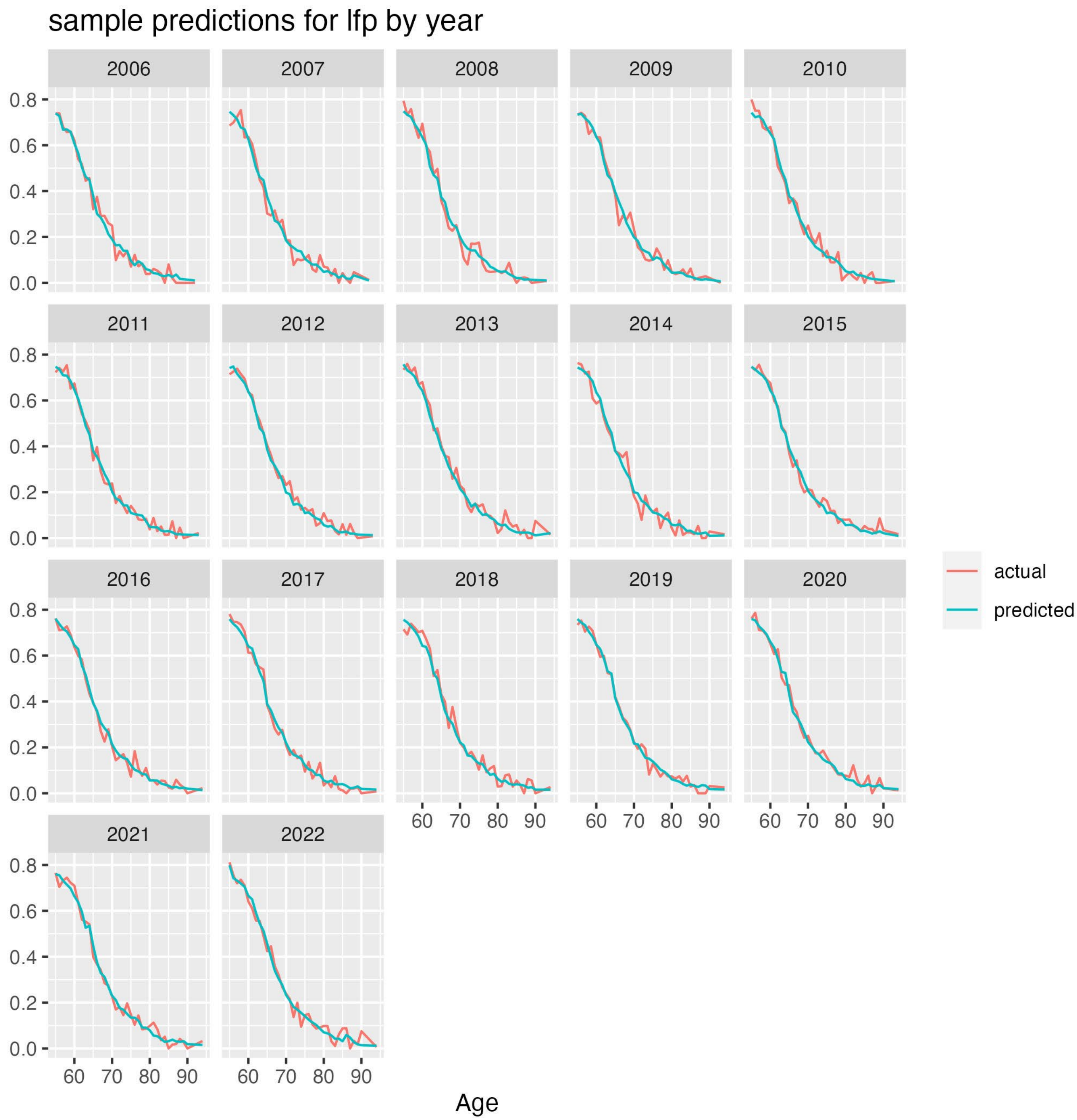
Predicted vs. actual from 80% training sample: poor (less than twice poverty level) by race, women only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B22

Predicted vs. actual from 80% training sample: labor force participation by year

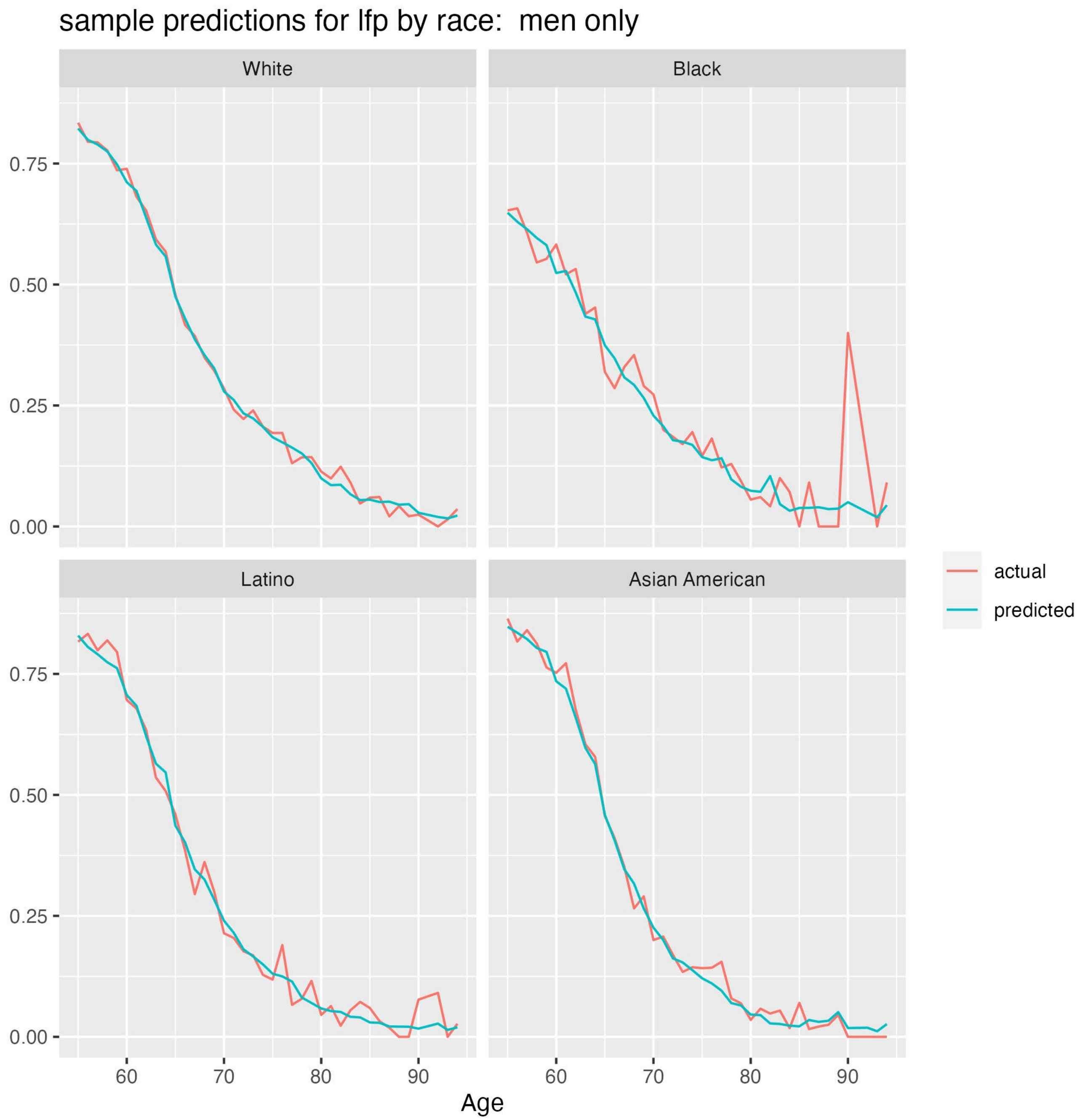


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B23

Predicted vs. actual from 80% training sample: labor force participation by race, men only

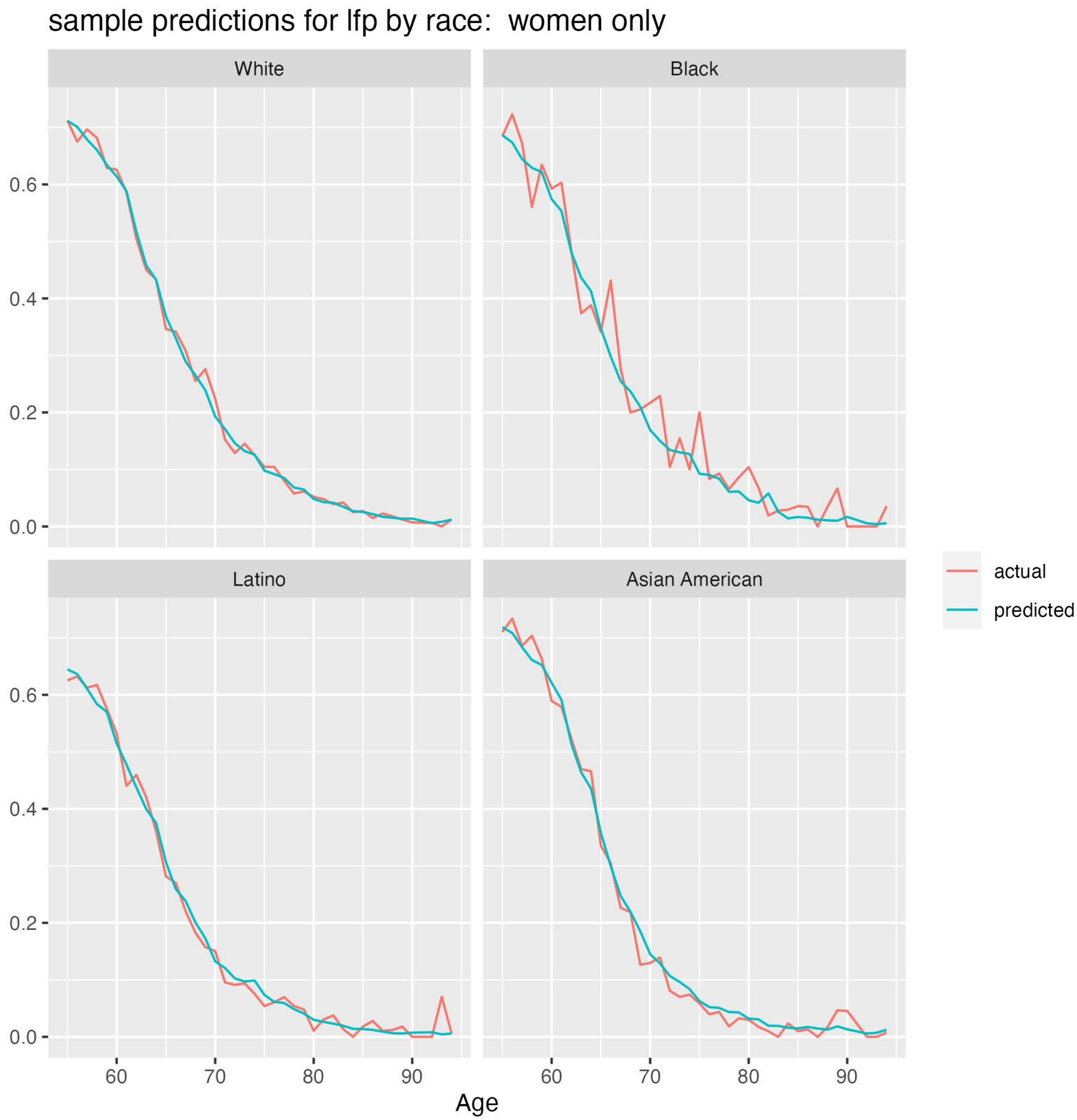


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B24

Predicted vs. actual from 80% training sample: labor force participation by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B25

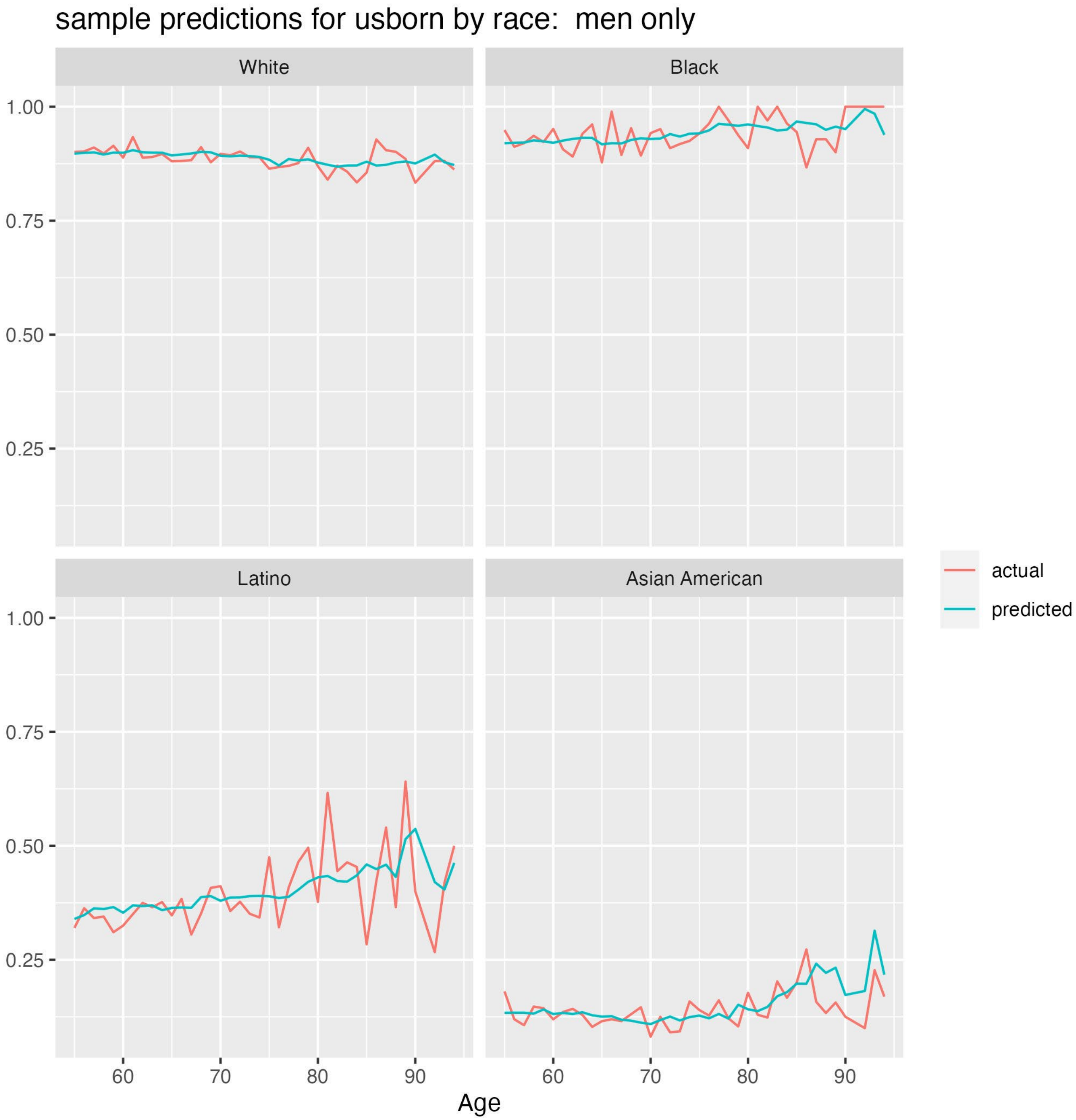
Predicted vs. actual from 80% training sample: U.S. born by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B26

Predicted vs. actual from 80% training sample: U.S. born by race, men only

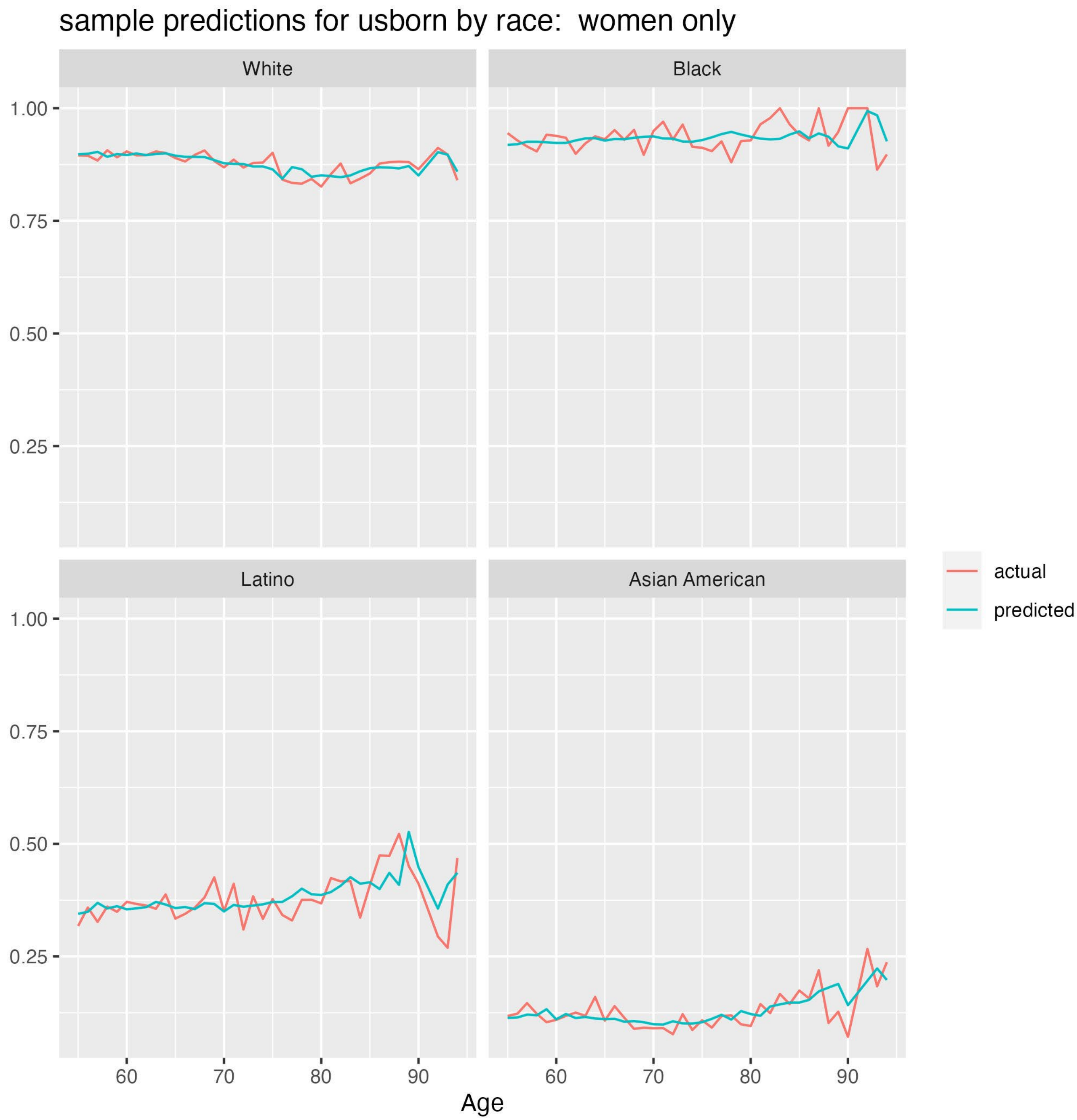


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B27

Predicted vs. actual from 80% training sample: U.S. born by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B28

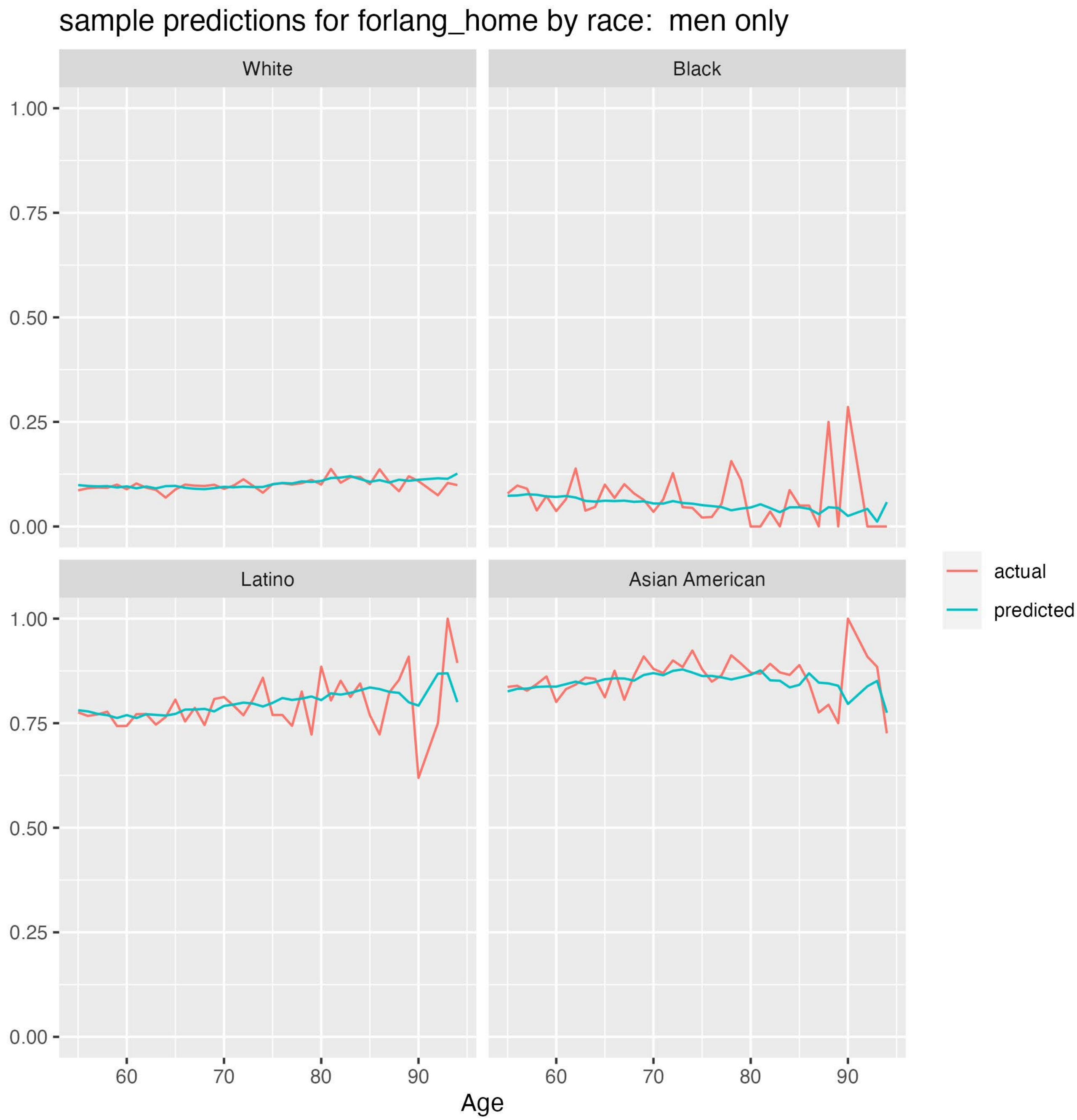
Predicted vs. actual from 80% training sample: speak foreign language at home by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B29

Predicted vs. actual from 80% training sample: speak foreign language at home by race, men only

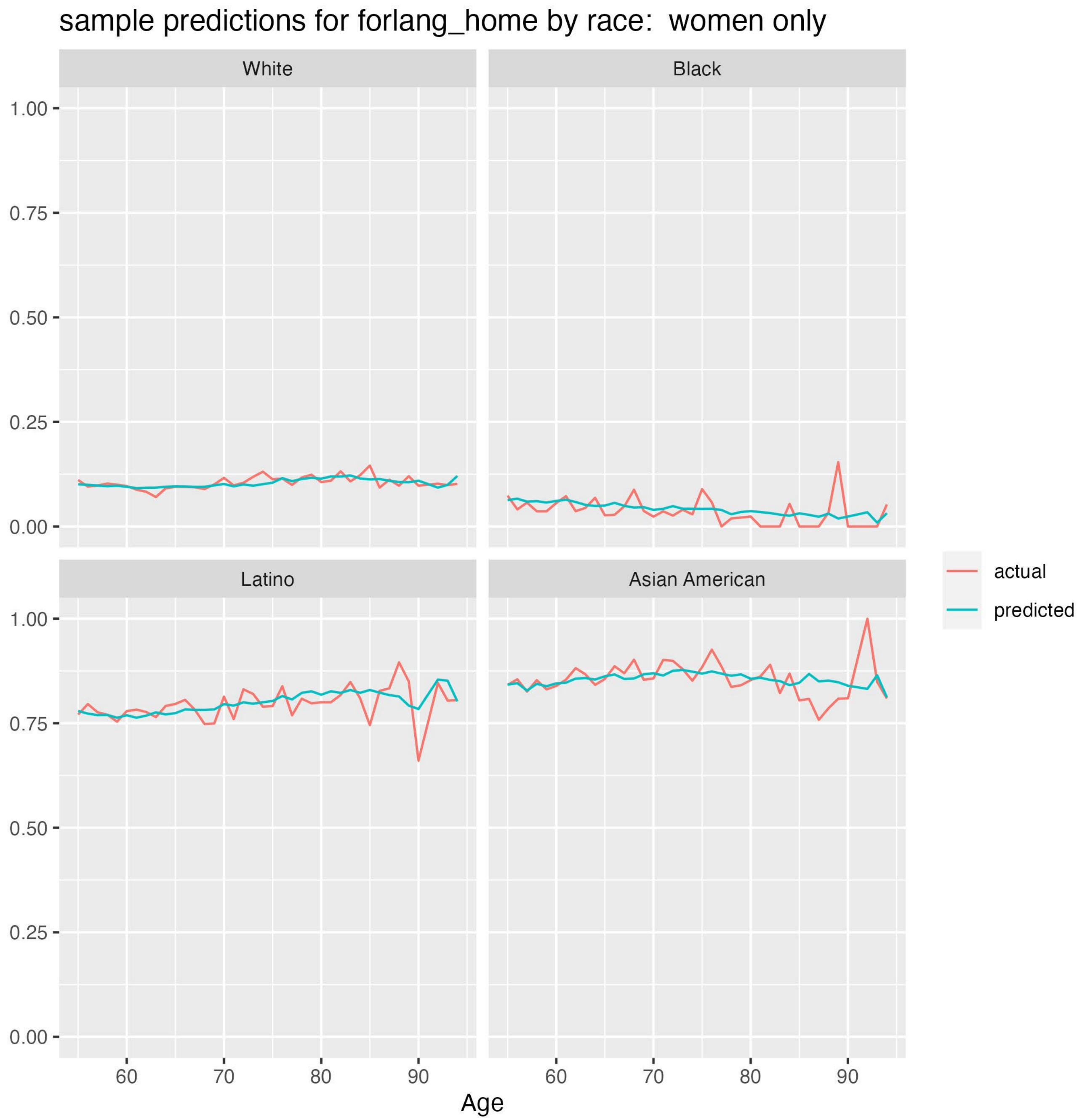


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B30

Predicted vs. actual from 80% training sample: speak foreign language at home by race, women only

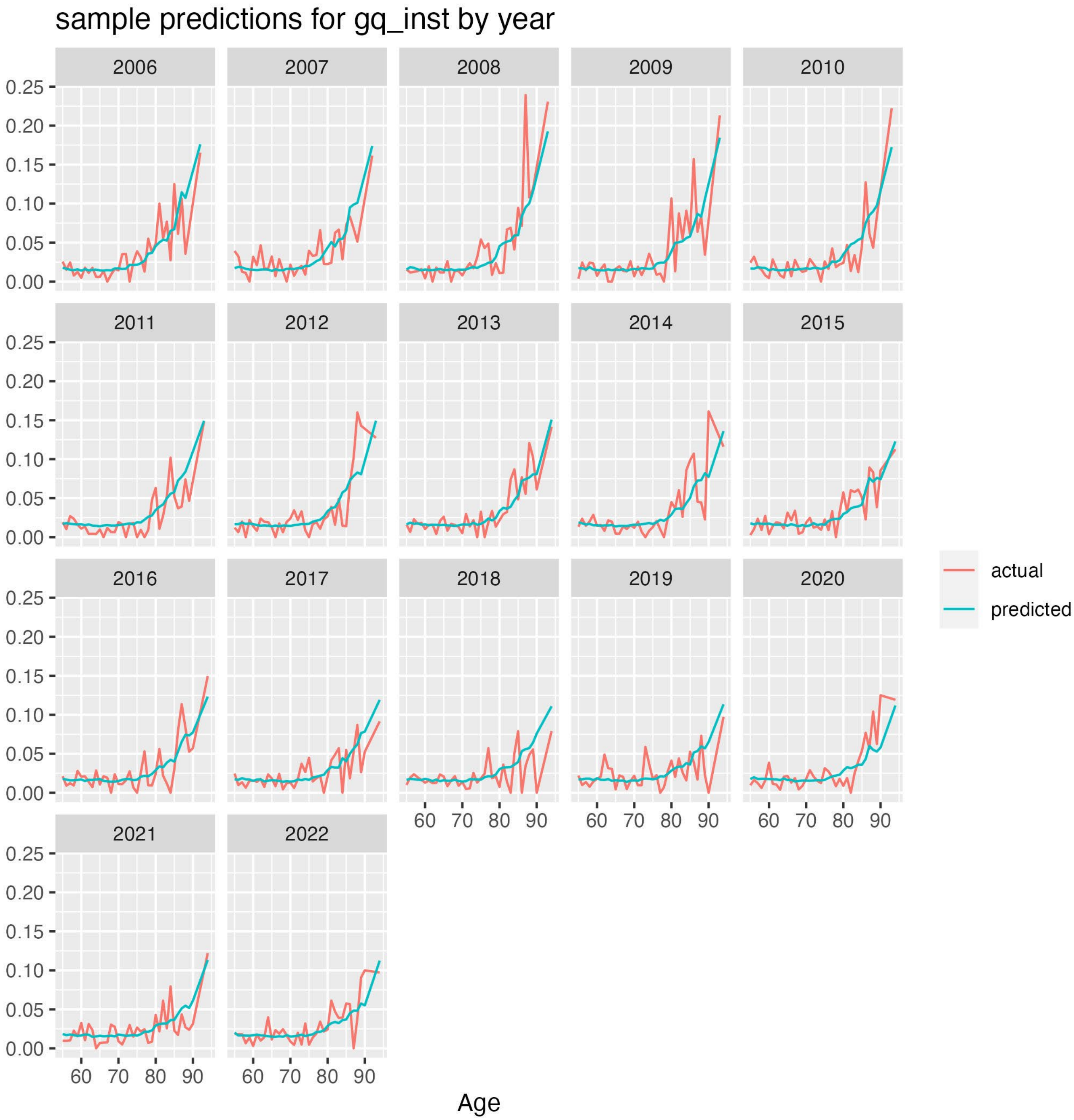


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B31

Predicted vs. actual from 80% training sample: live in institutional group quarters by year

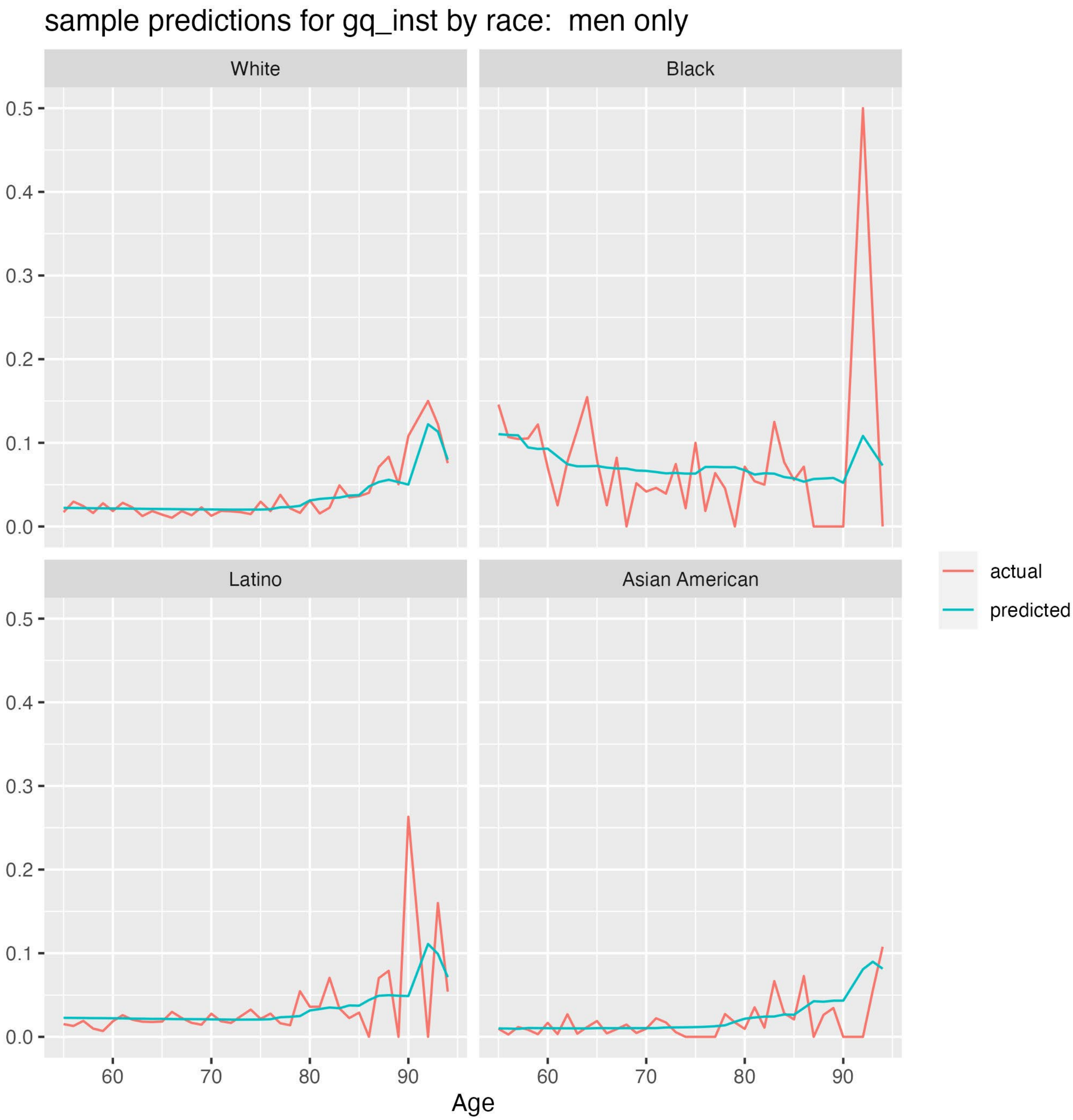


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B32

Predicted vs. actual from 80% training sample: live in institutional group quarters by race, men only

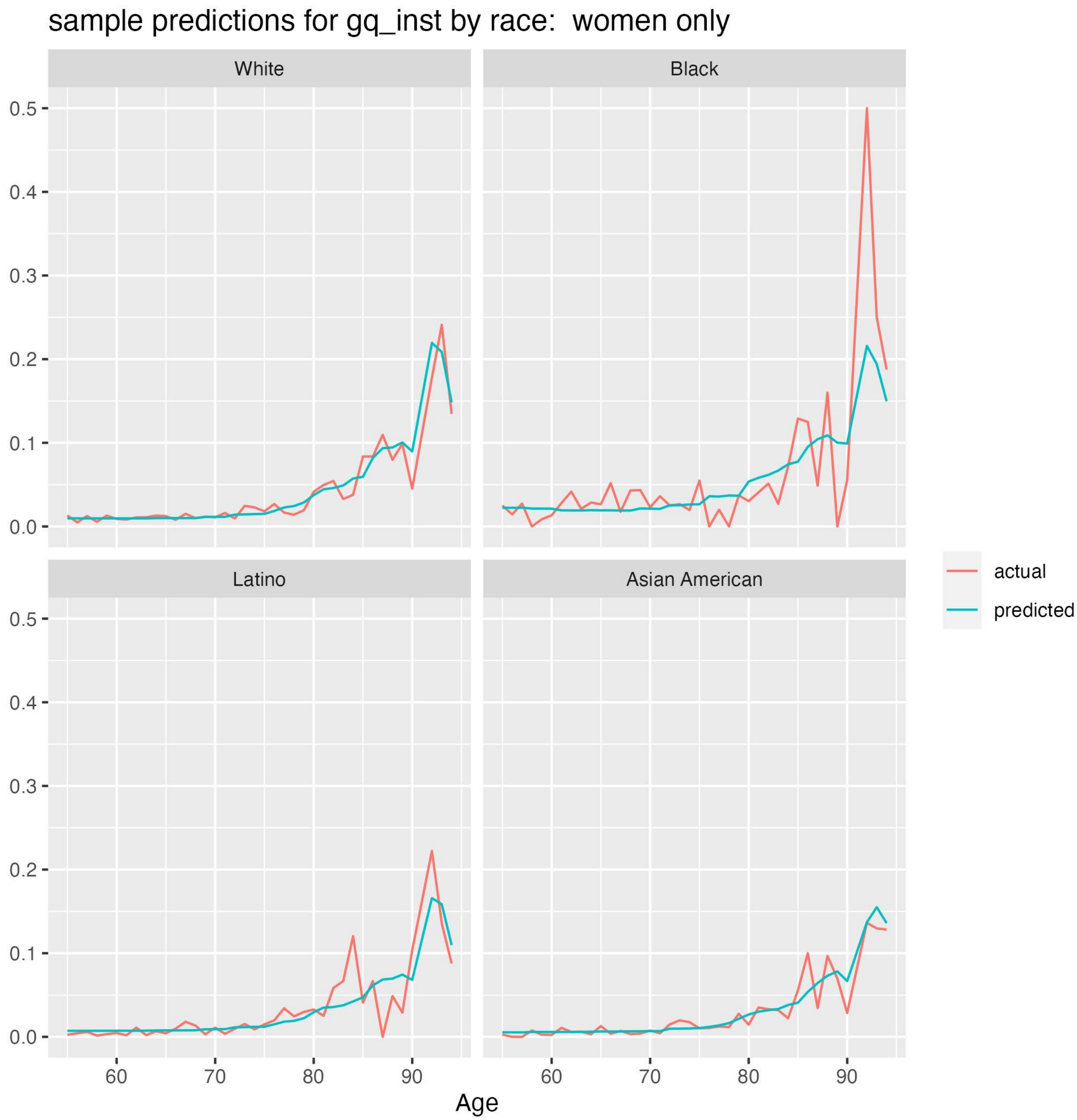


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B33

Predicted vs. actual from 80% training sample: live in institutional group quarters by race, women only

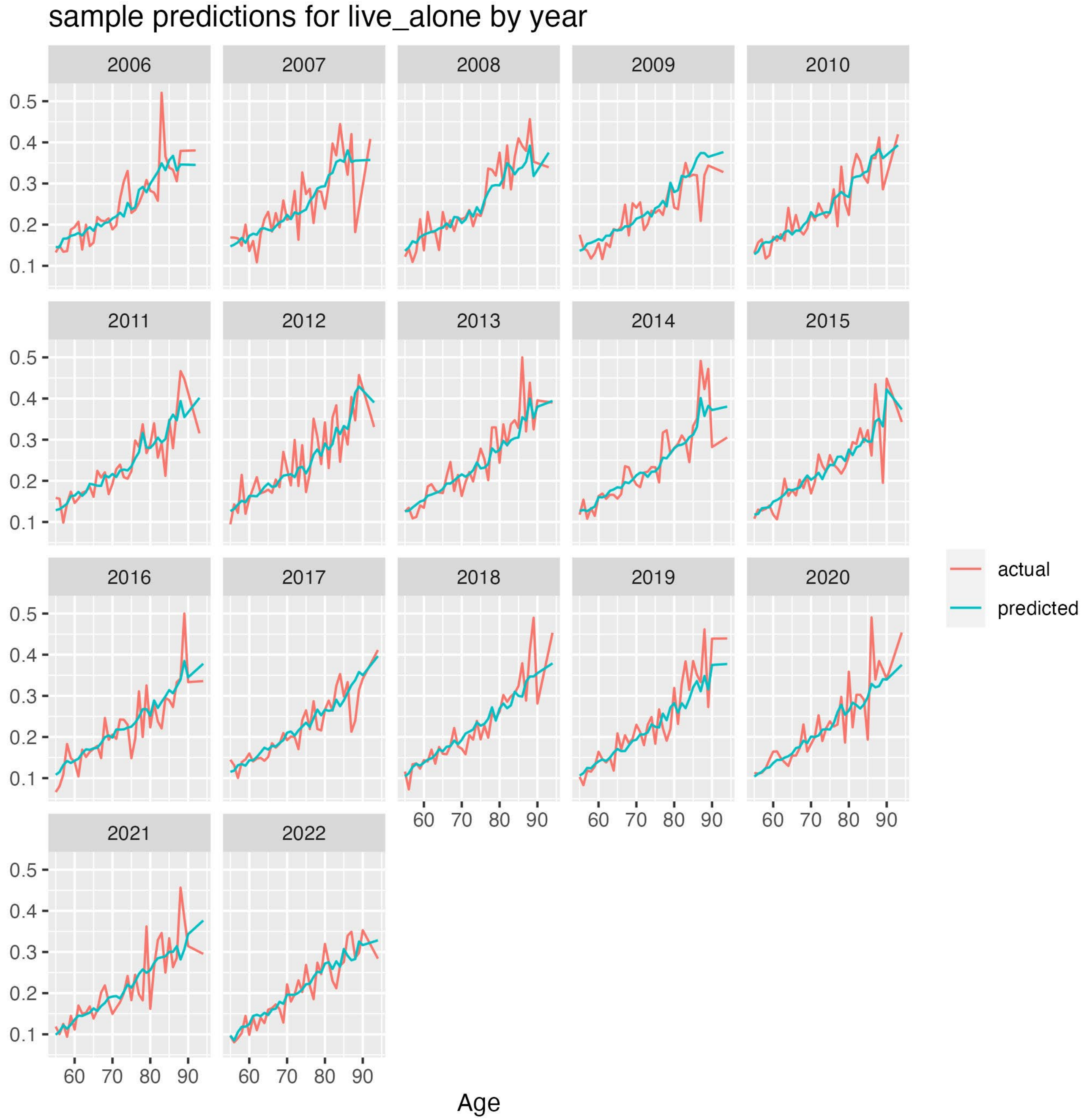


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B34

Predicted vs. actual from 80% training sample: live alone by year

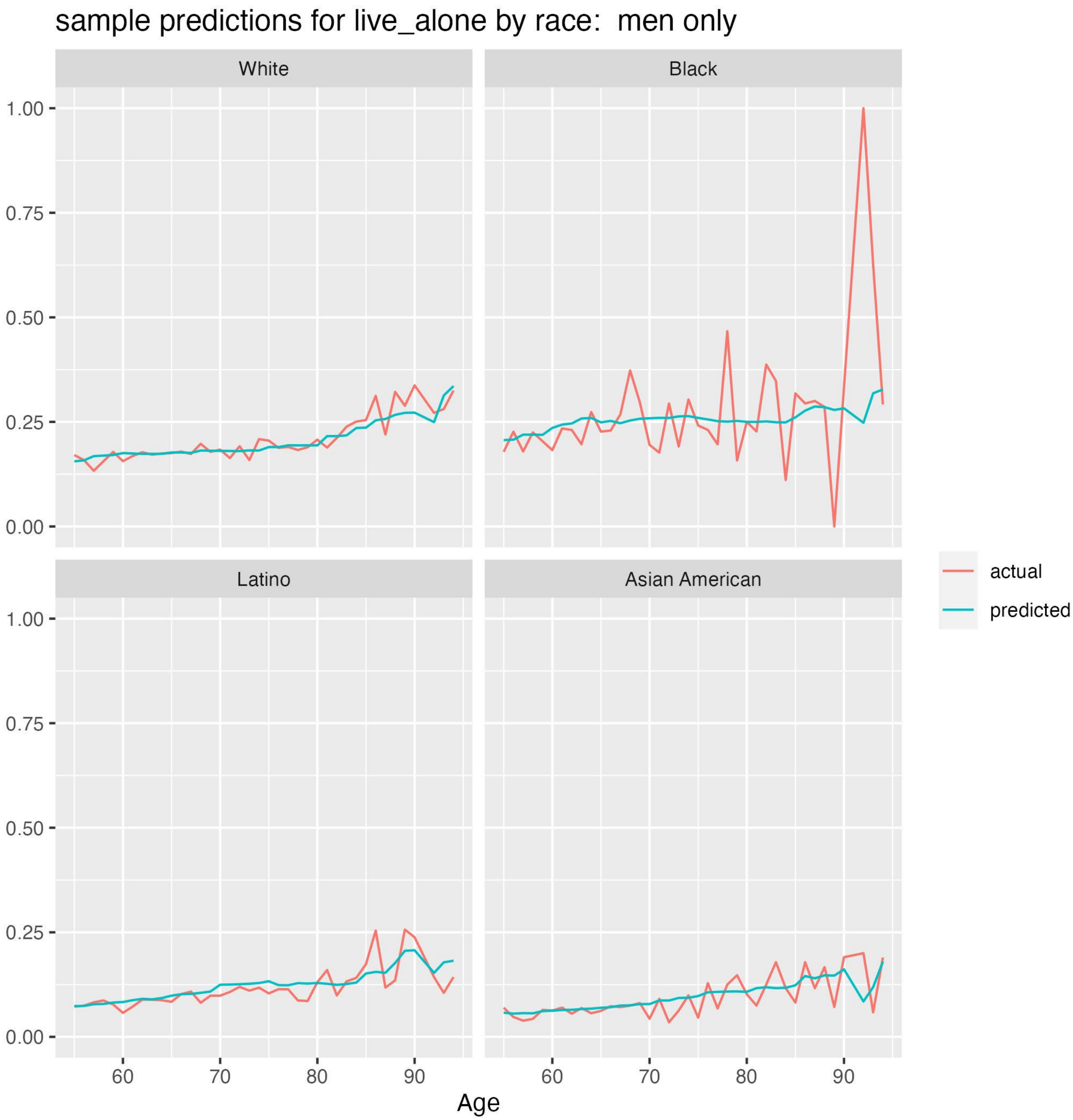


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B35

Predicted vs. actual from 80% training sample: live alone by race, men only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B36

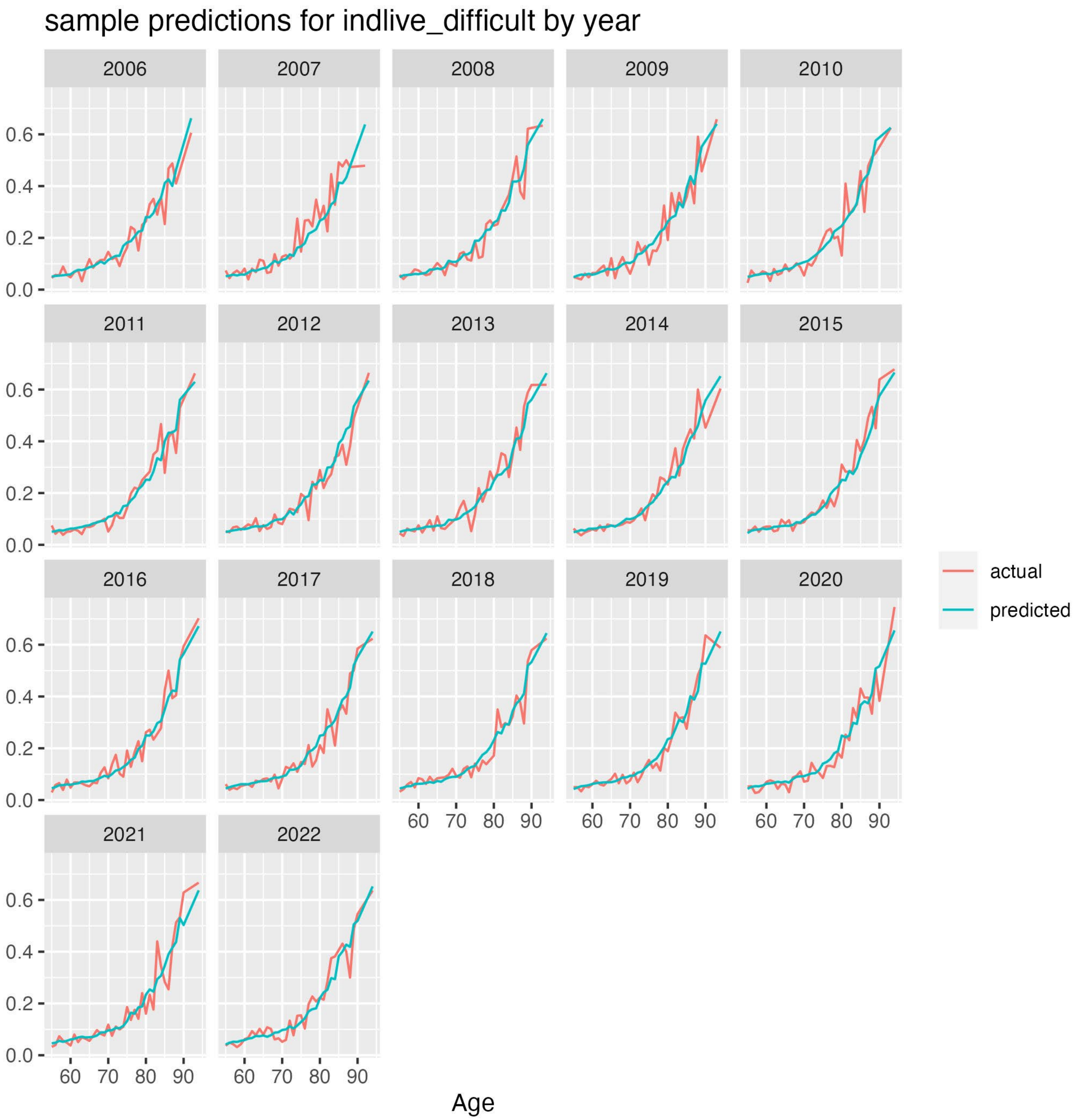
Predicted vs. actual from 80% training sample: live alone by race, women only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B37

Predicted vs. actual from 80% training sample: difficulty living independently by year

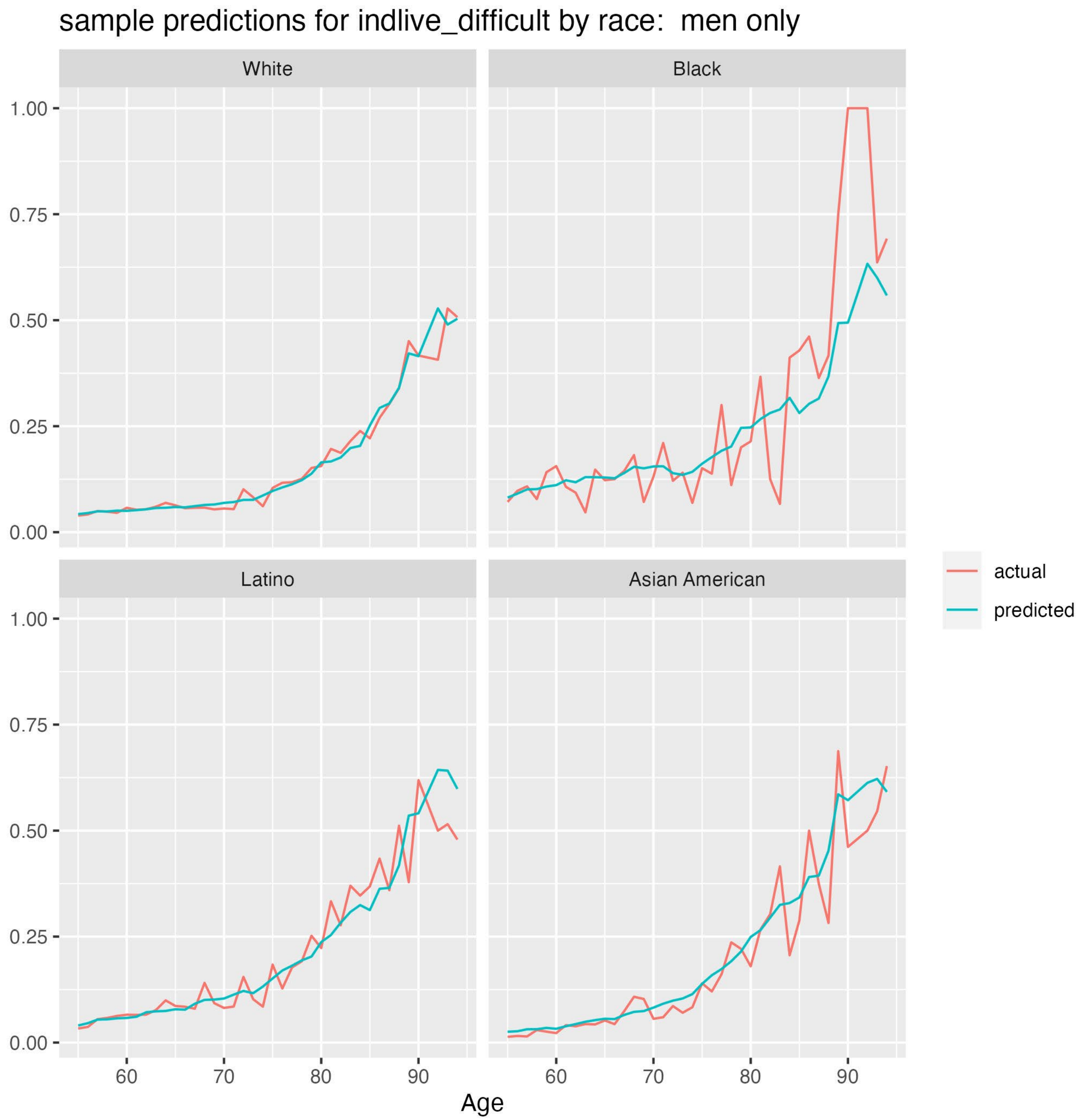


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B38

Predicted vs. actual from 80% training sample: difficulty living independently by race, men only

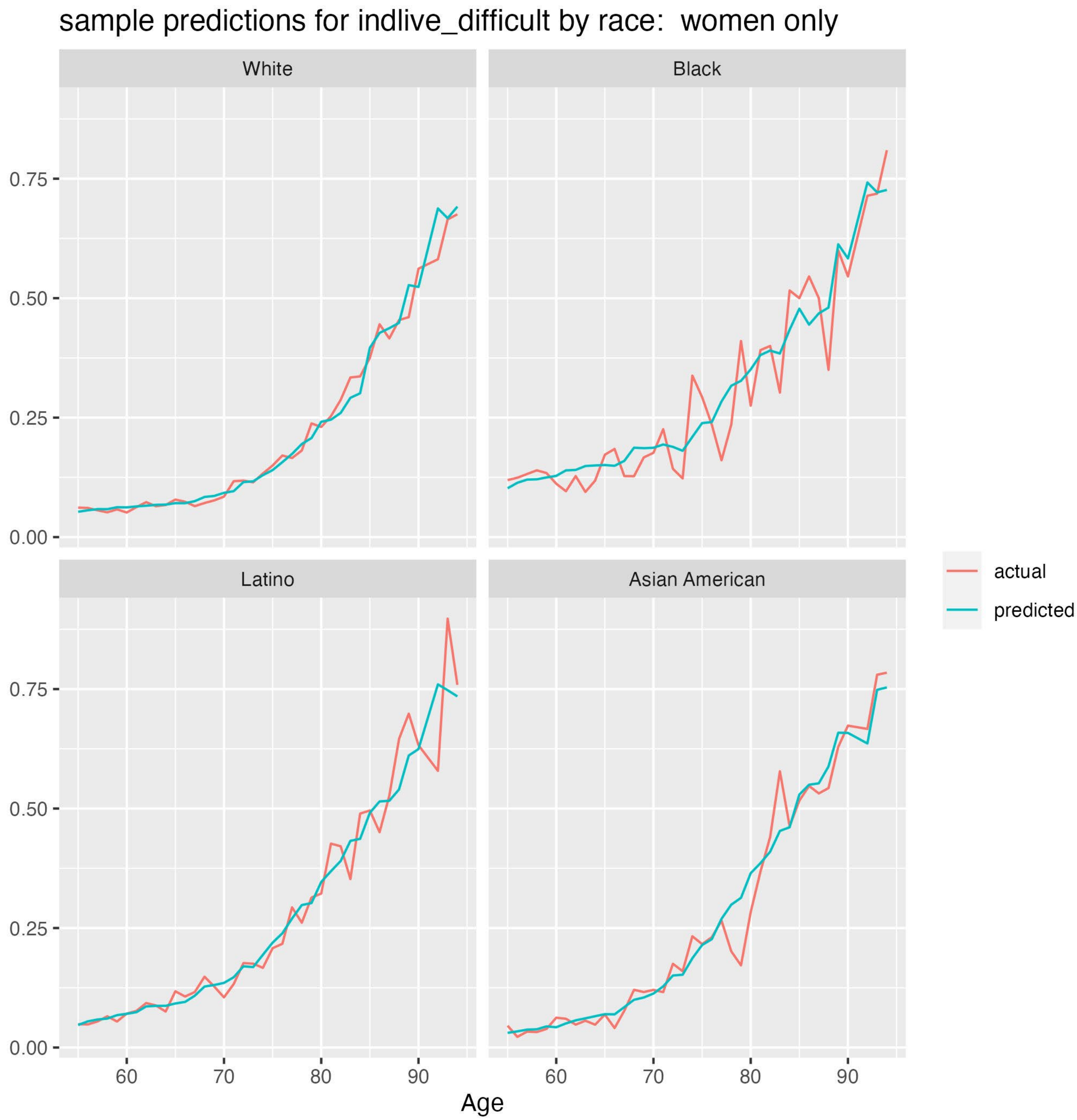


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B39

Predicted vs. actual from 80% training sample: difficulty living independently by race, women only

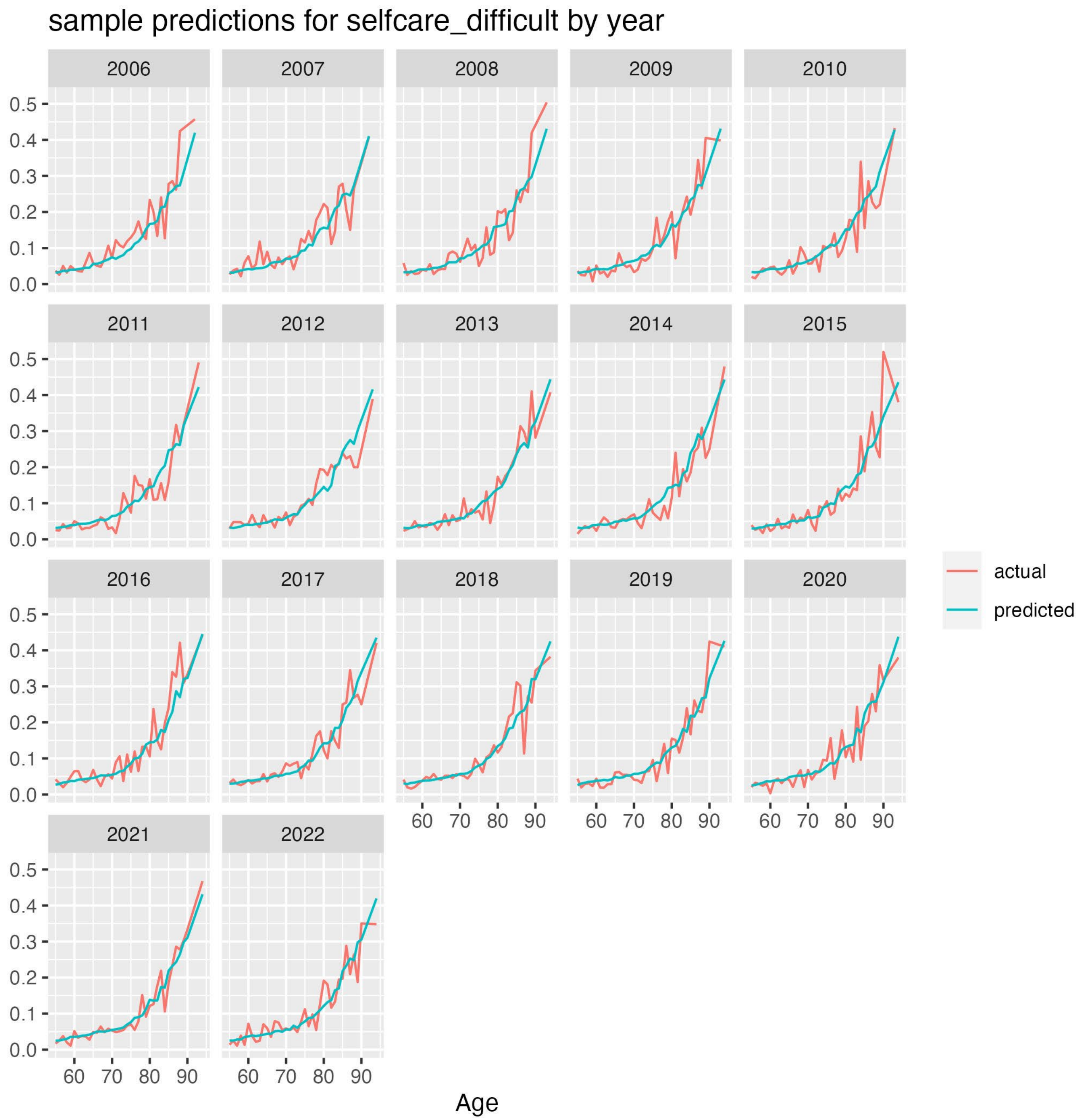


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B40

Predicted vs. actual from 80% training sample: difficulty with self care by year

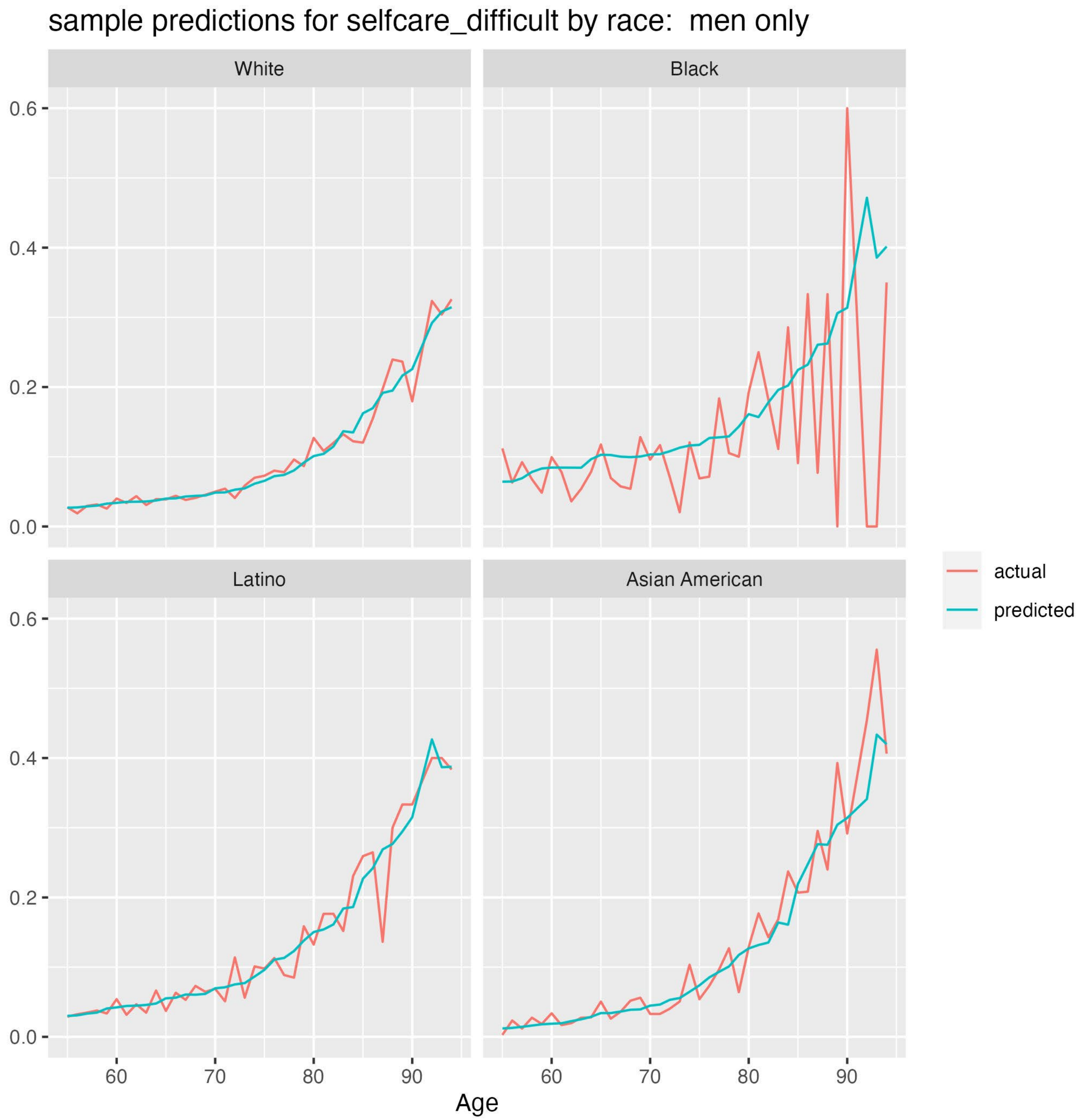


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B41

Predicted vs. actual from 80% training sample: difficulty with self care by race, men only

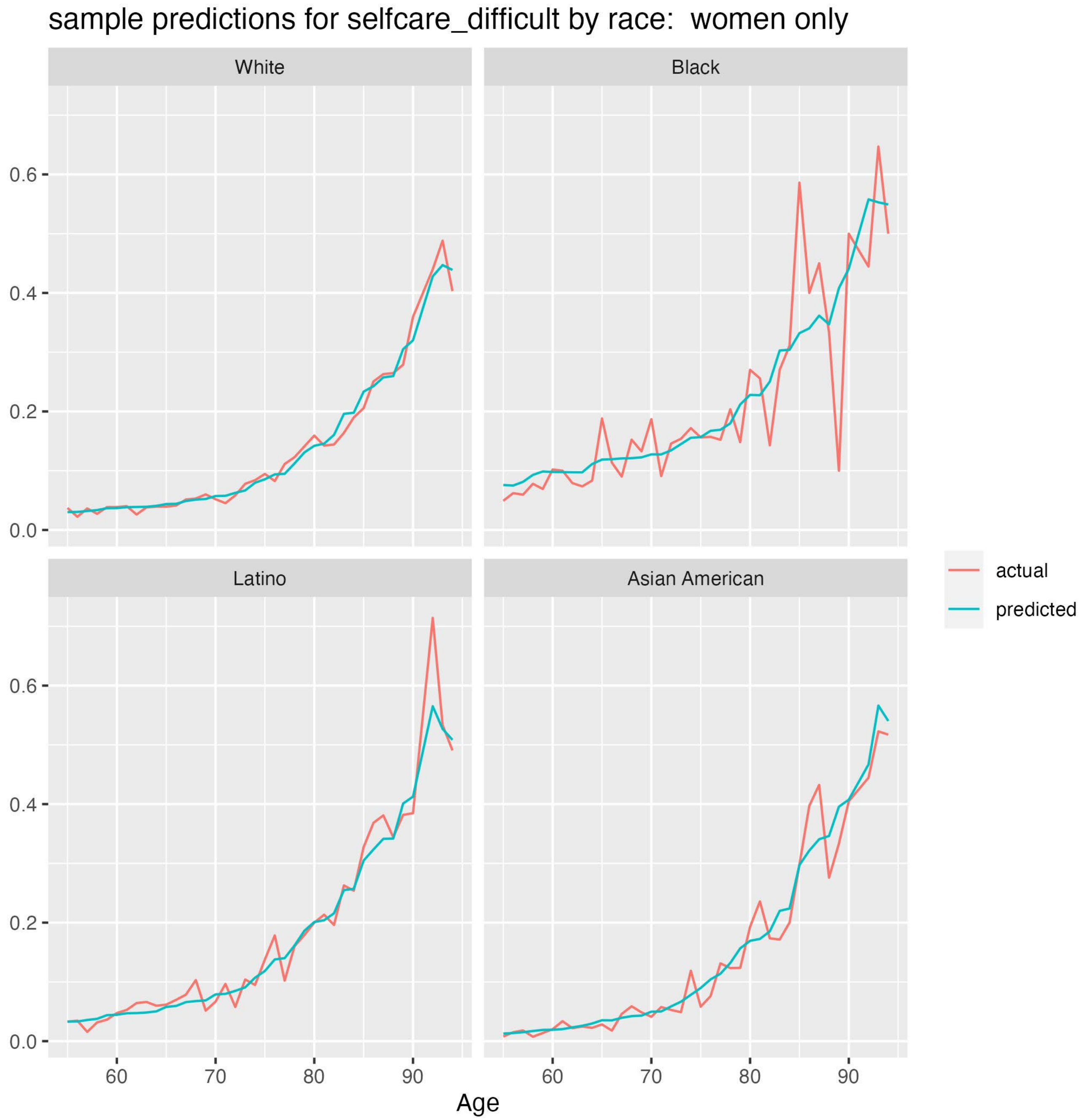


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B42

Predicted vs. actual from 80% training sample: difficulty with self care by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on a random subset of 80% of the data, and tested on the remaining 20%.

FIGURE B43

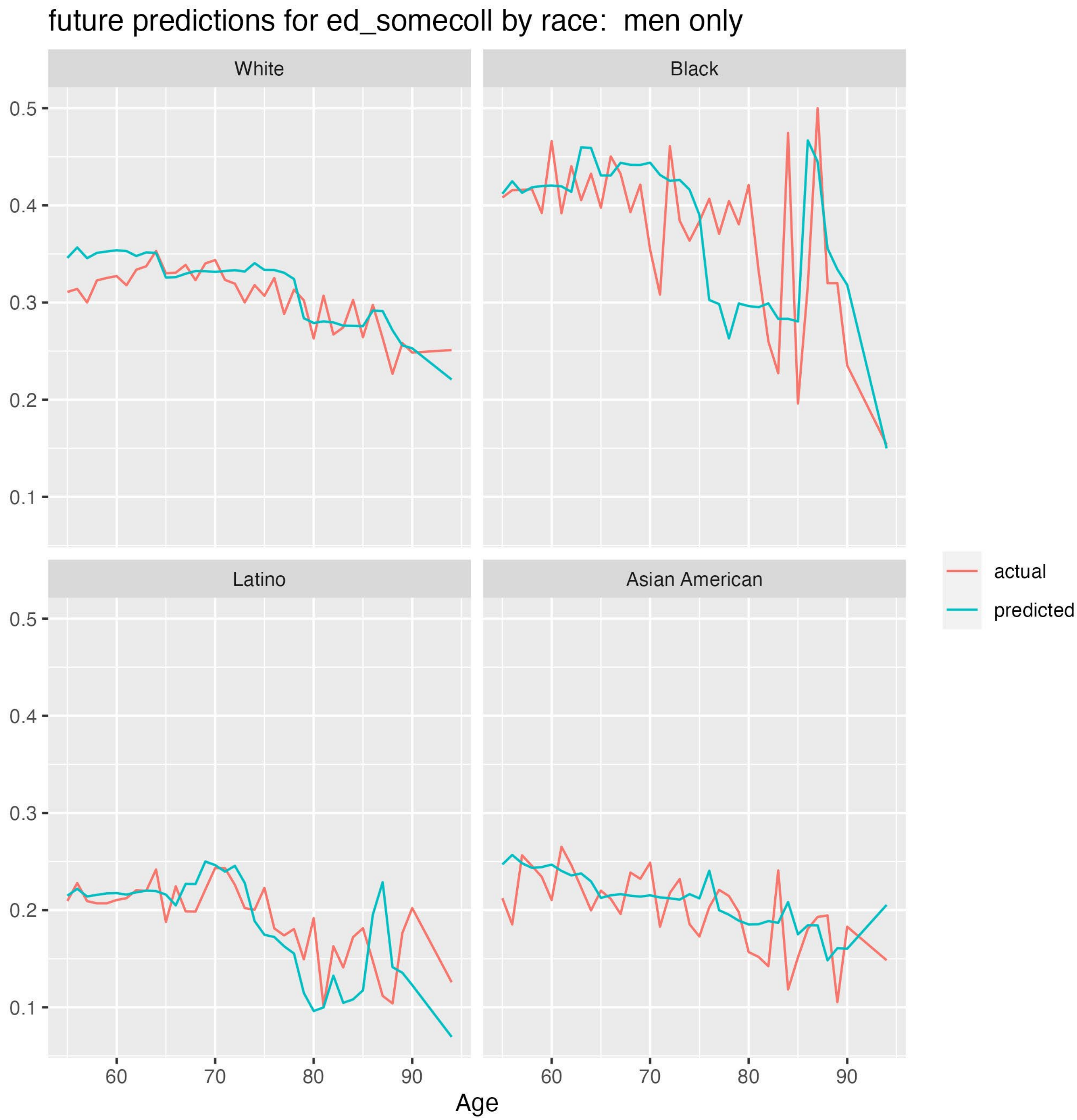
Predicted vs. actual from time split: college graduates by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B44

Predicted vs. actual from time split: college graduates by race, men only

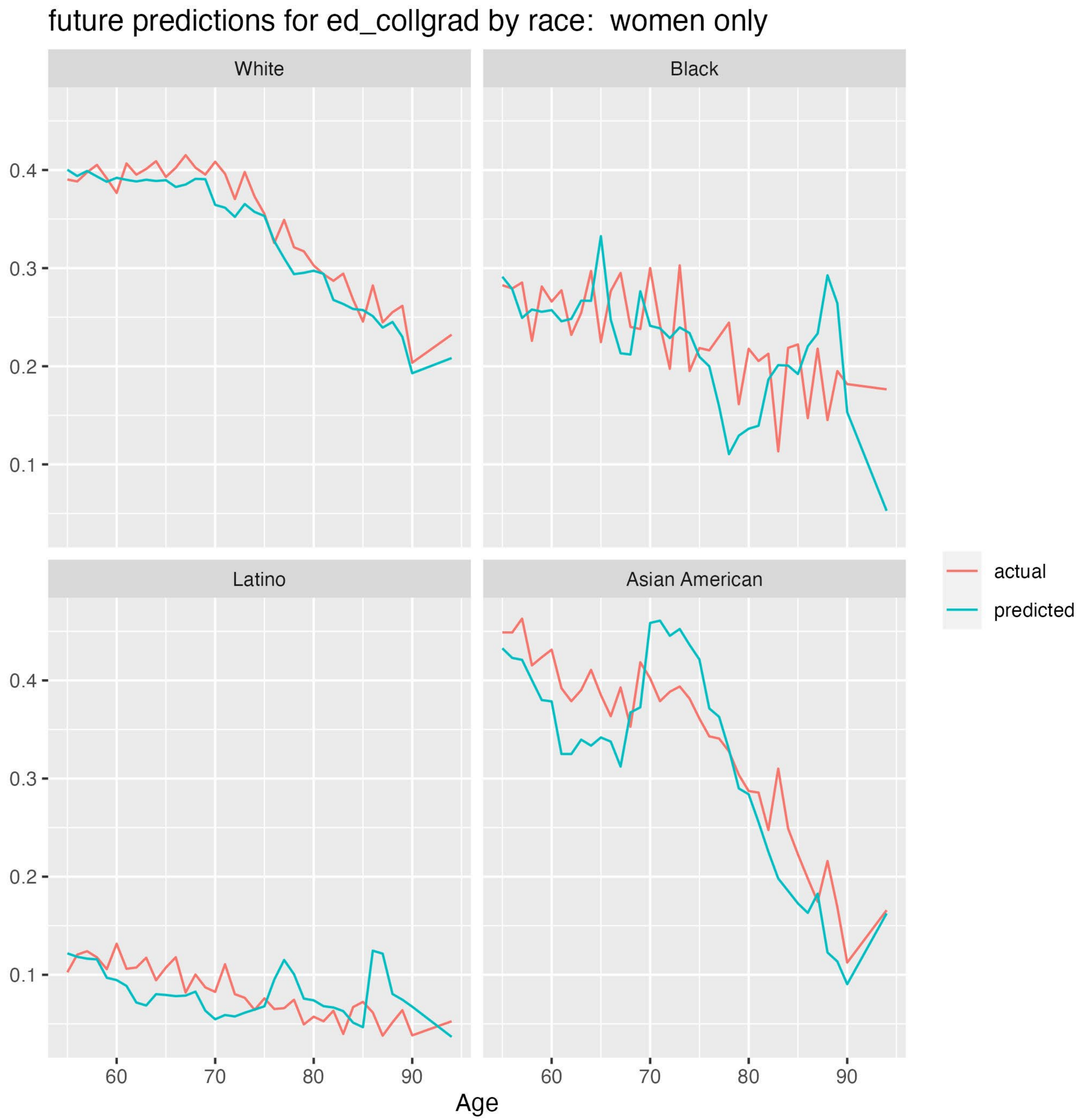


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B45

Predicted vs. actual from time split: college graduates by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B46

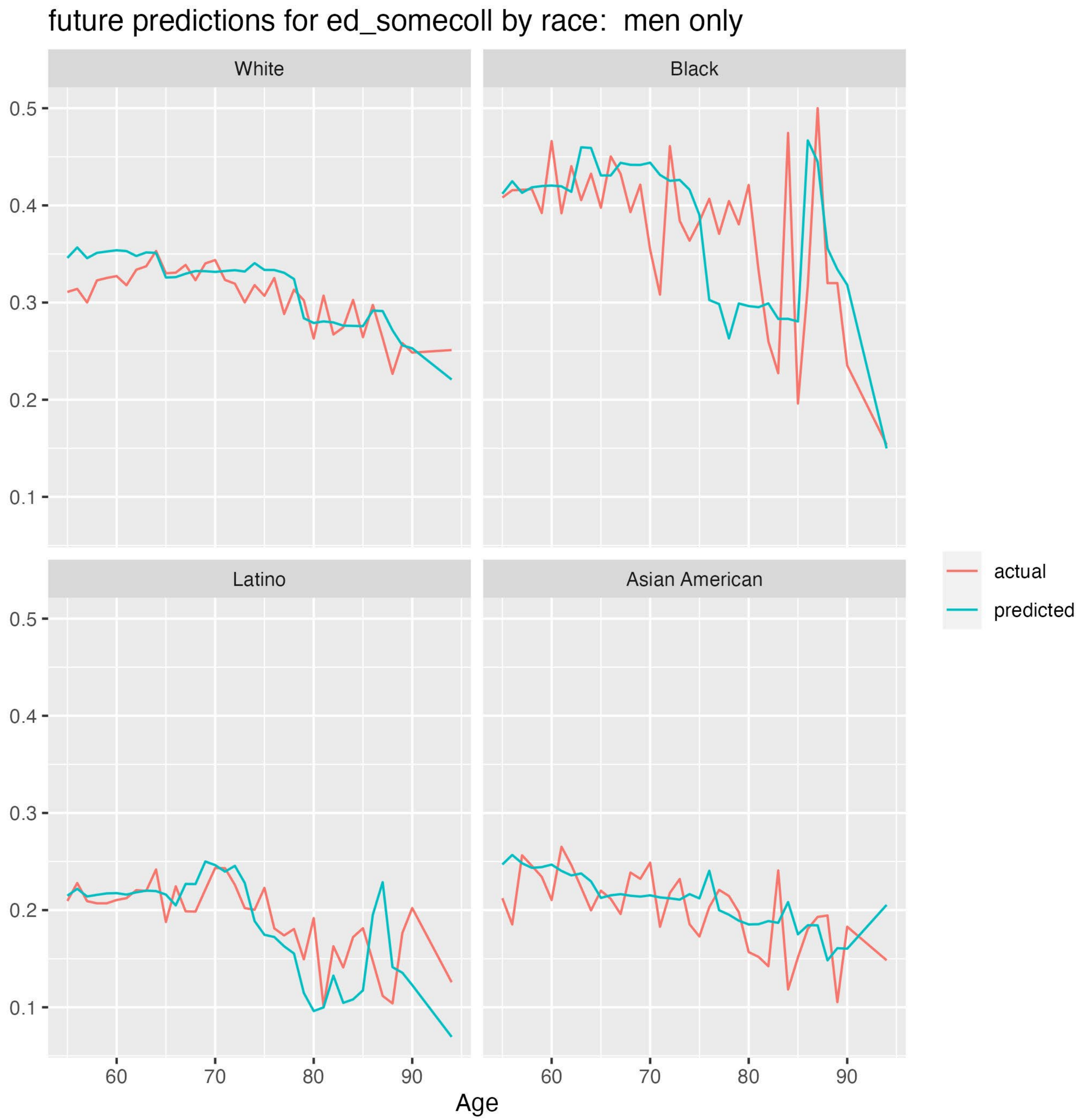
Predicted vs. actual from time split: some college by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B47

Predicted vs. actual from time split: some college by race, men only

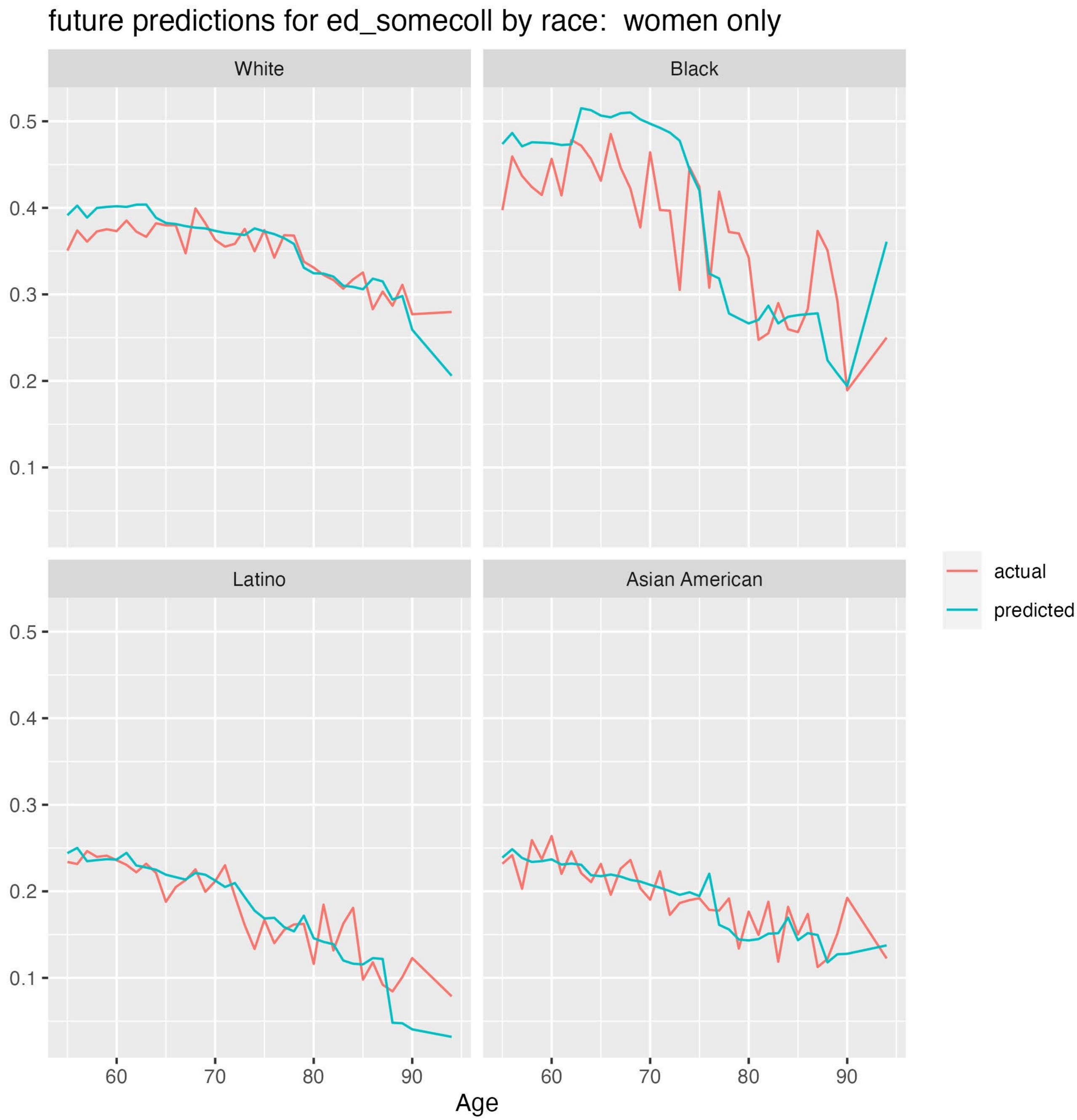


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B48

Predicted vs. actual from time split: some college by race, women only

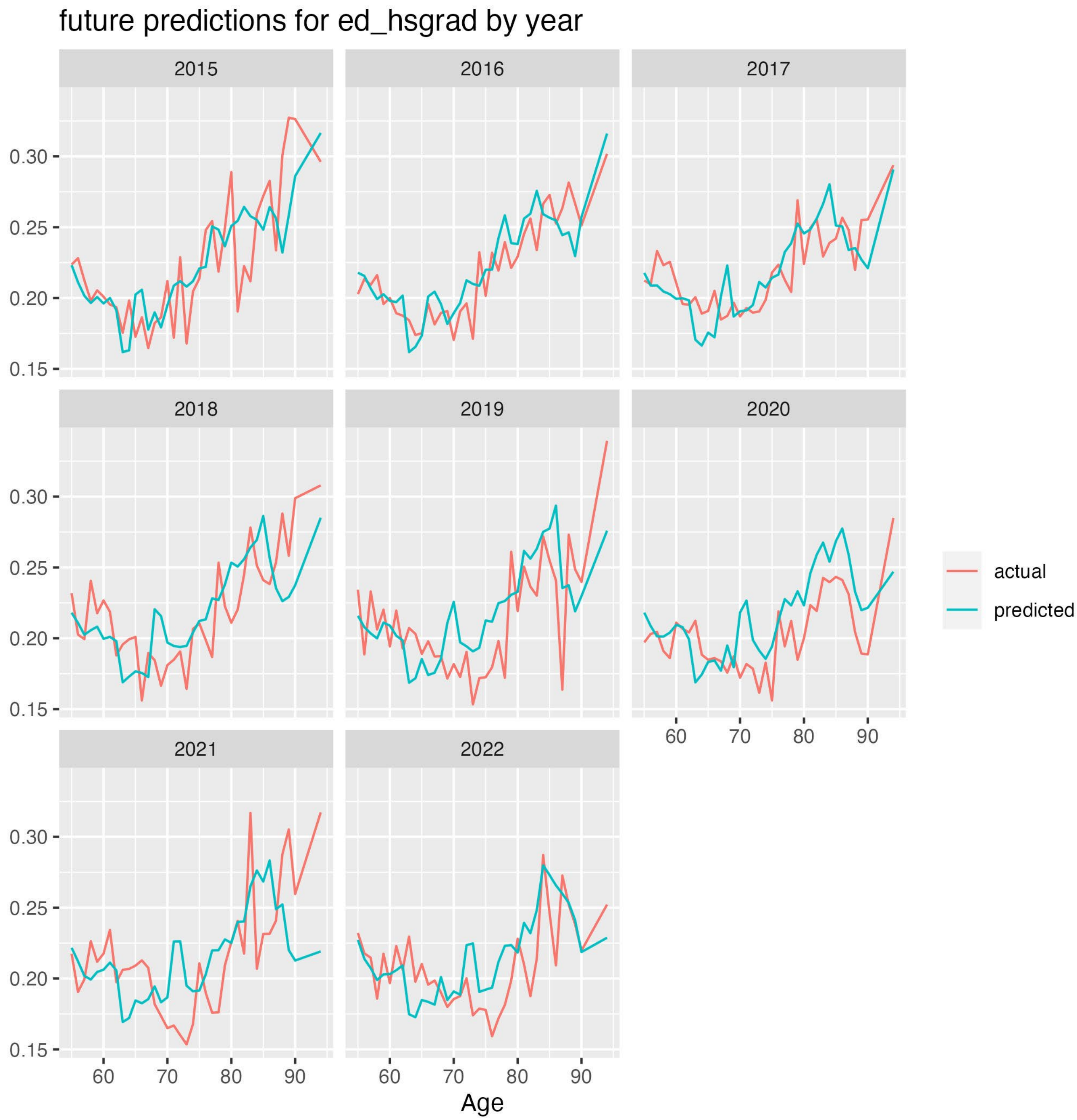


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B49

Predicted vs. actual from time split: high school graduates by year

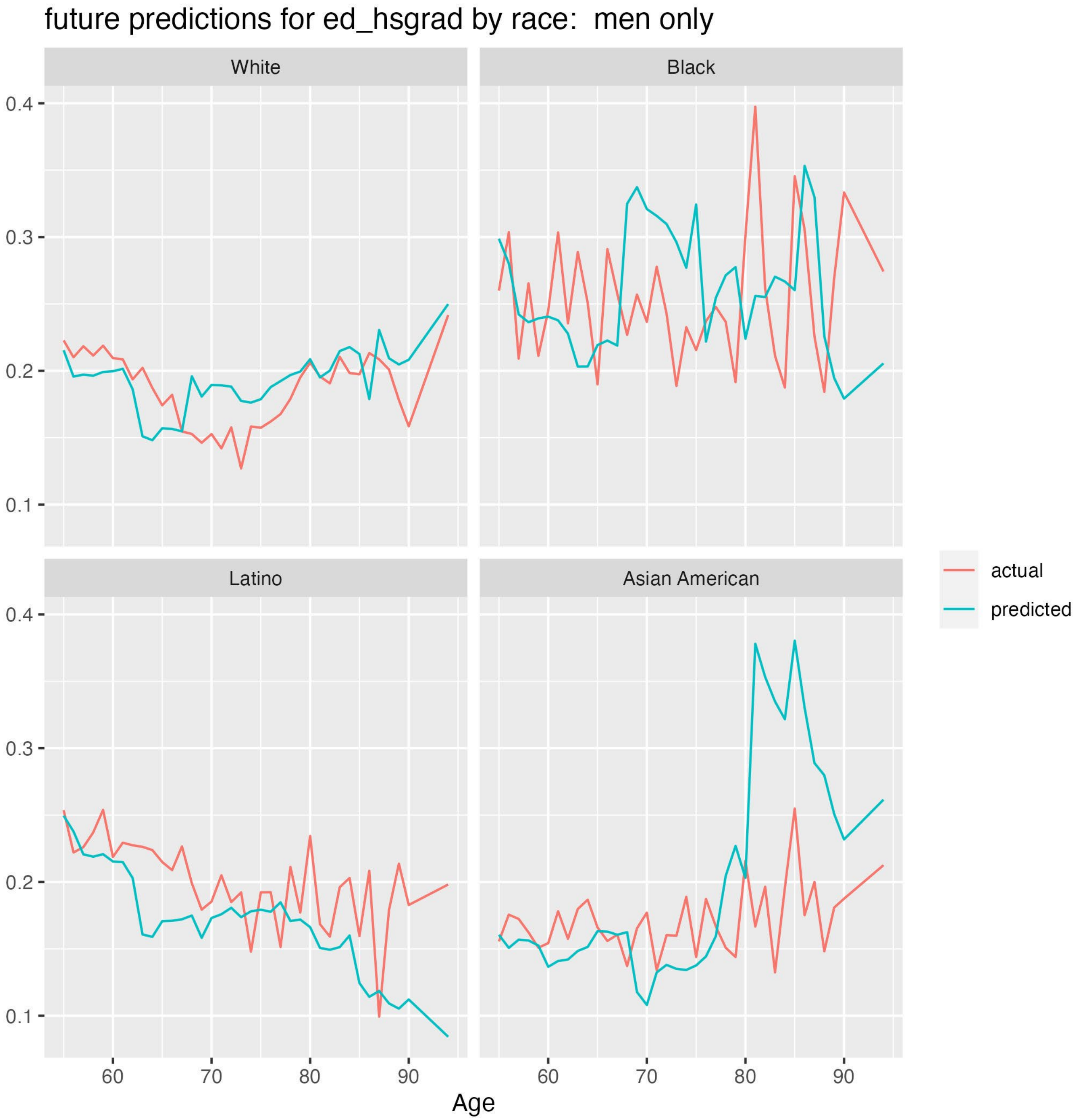


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B50

Predicted vs. actual from time split: high school graduates by race, men only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

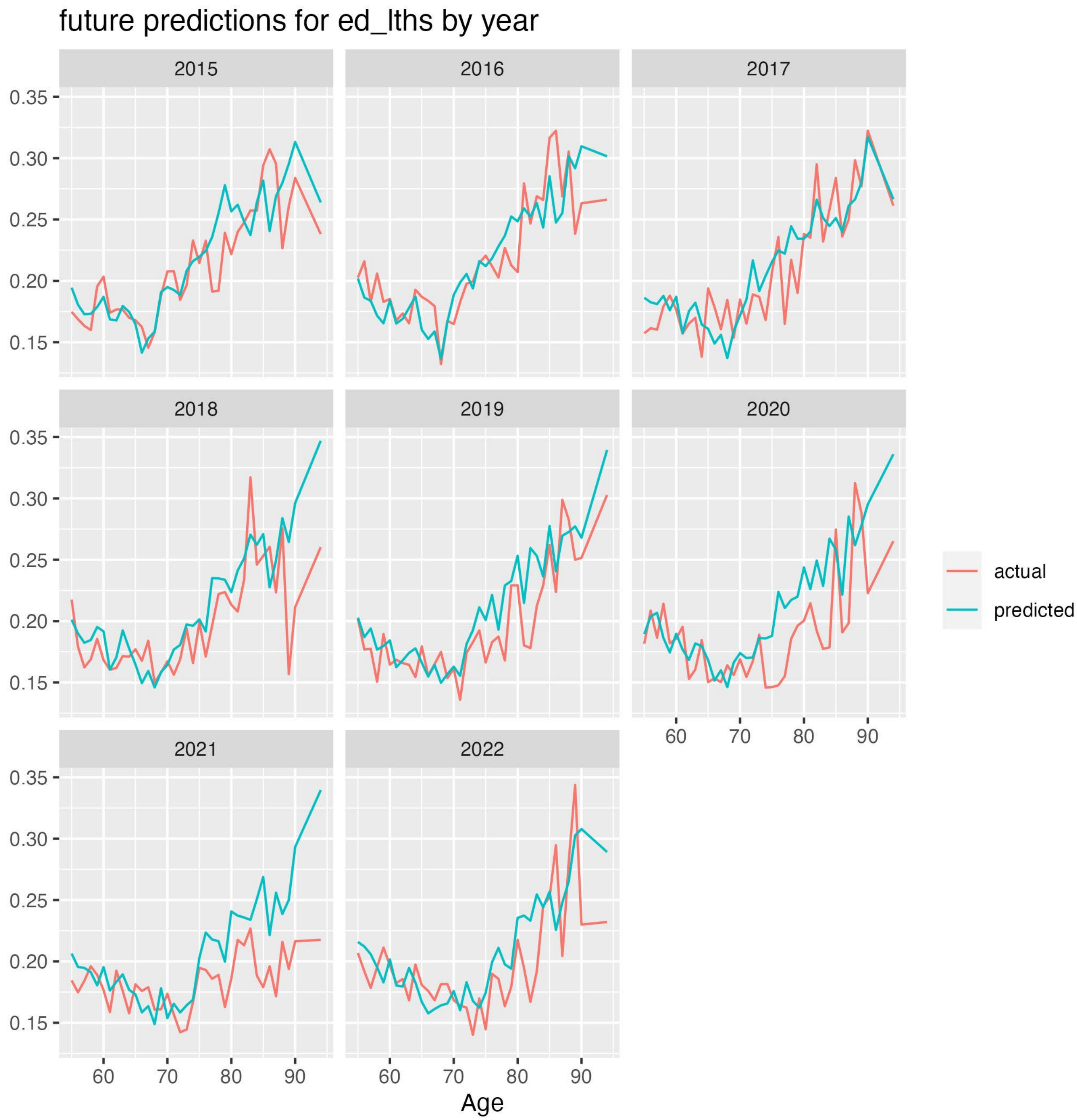
FIGURE B51

Predicted vs. actual from time split: high school graduates by race, women only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

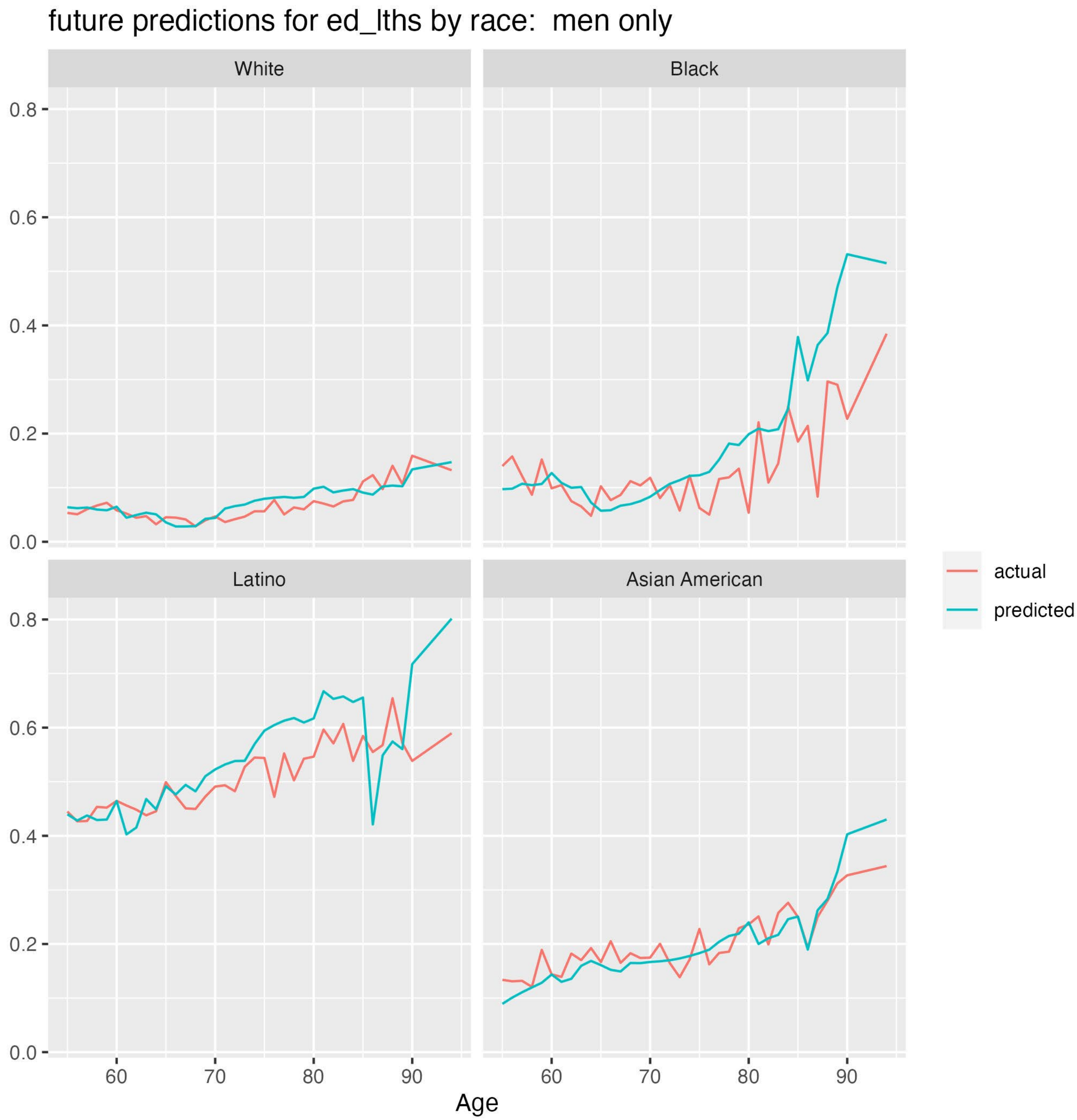
FIGURE B52
Predicted vs. actual from time split: less than high school by year



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.
NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B53

Predicted vs. actual from time split: less than high school by race, men only

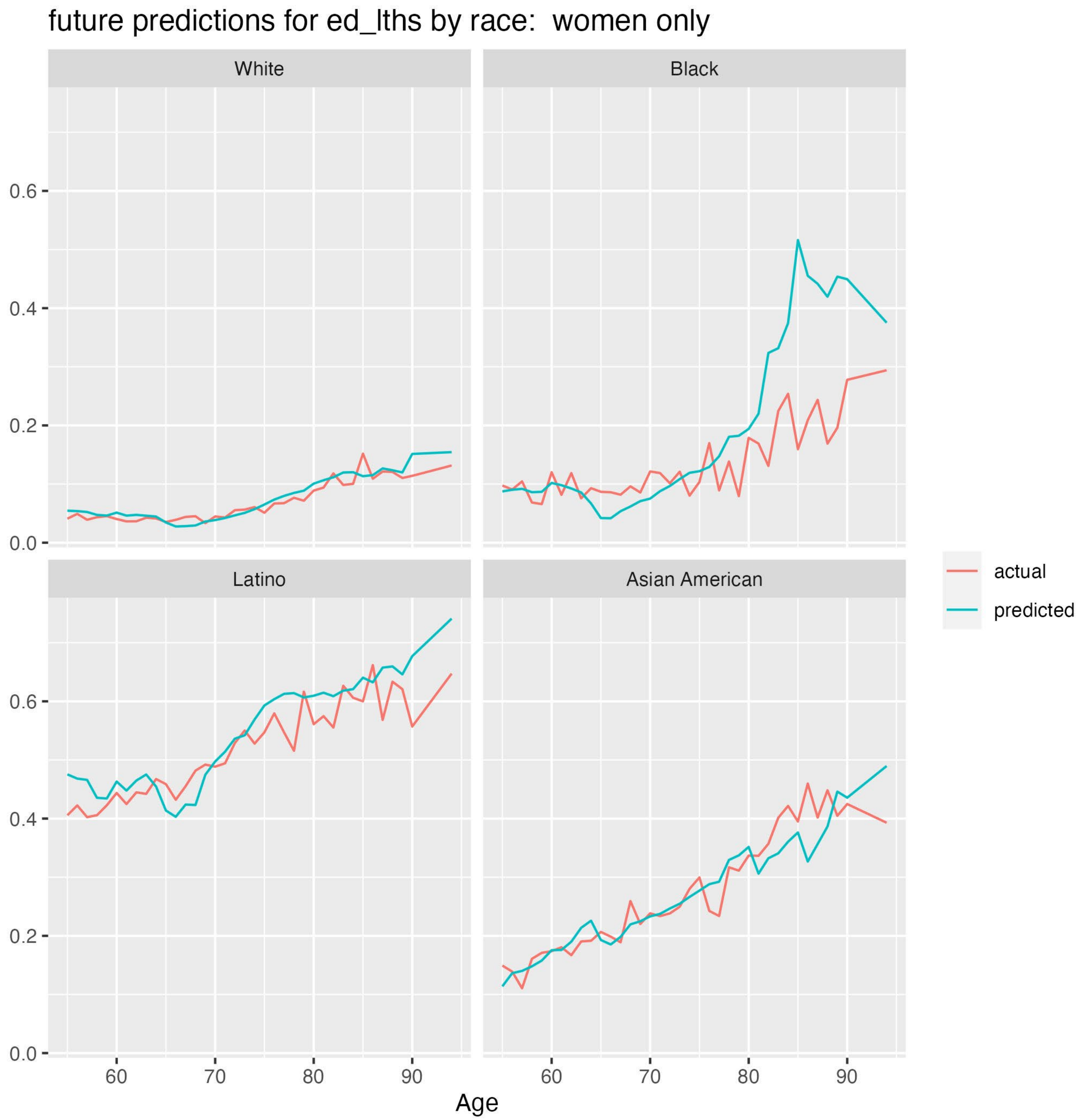


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B54

Predicted vs. actual from time split: less than high school by race, women only

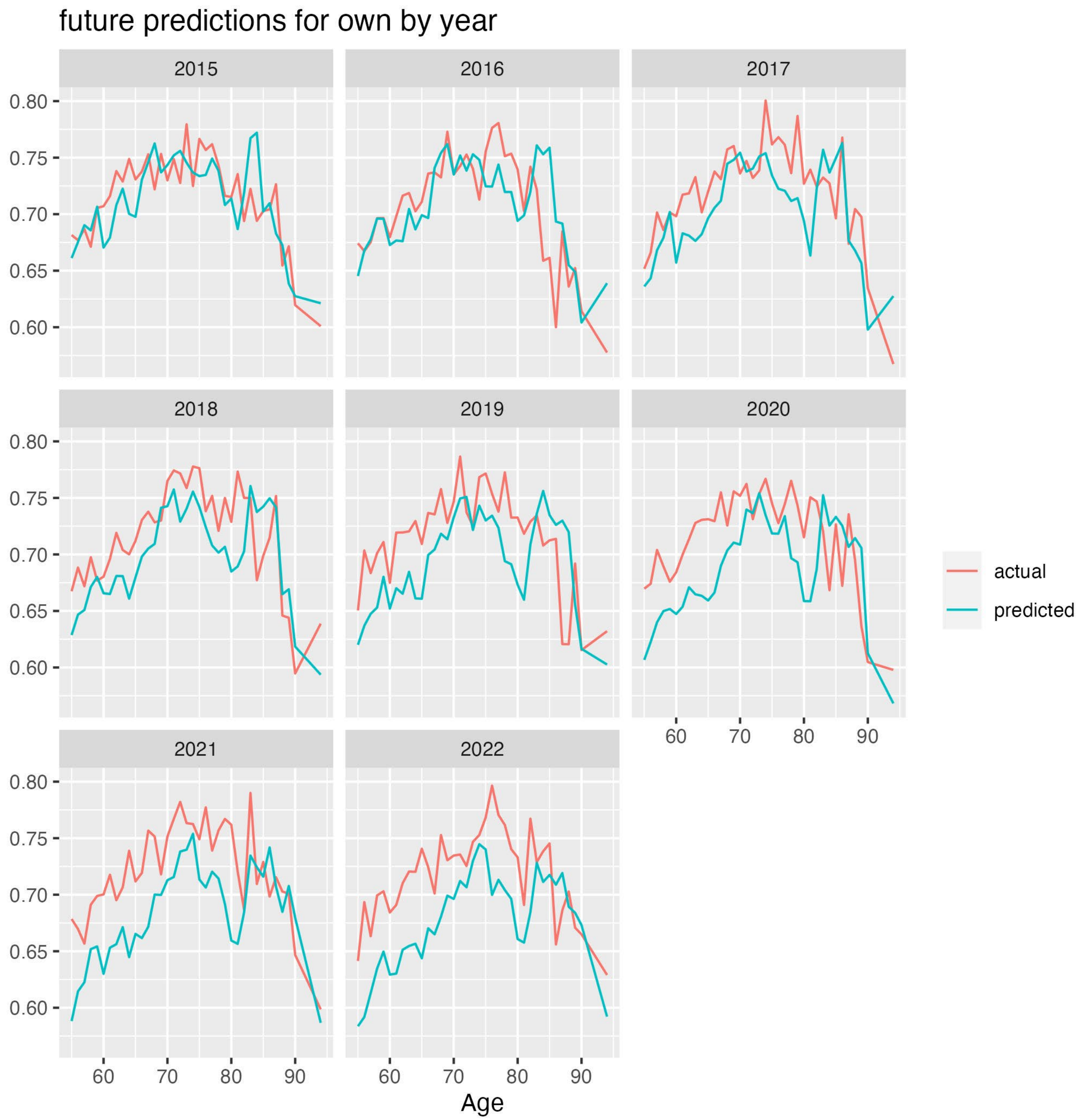


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B55

Predicted vs. actual from time split: home ownership by year



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B56

Predicted vs. actual from time split: home ownership by race, men only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B57

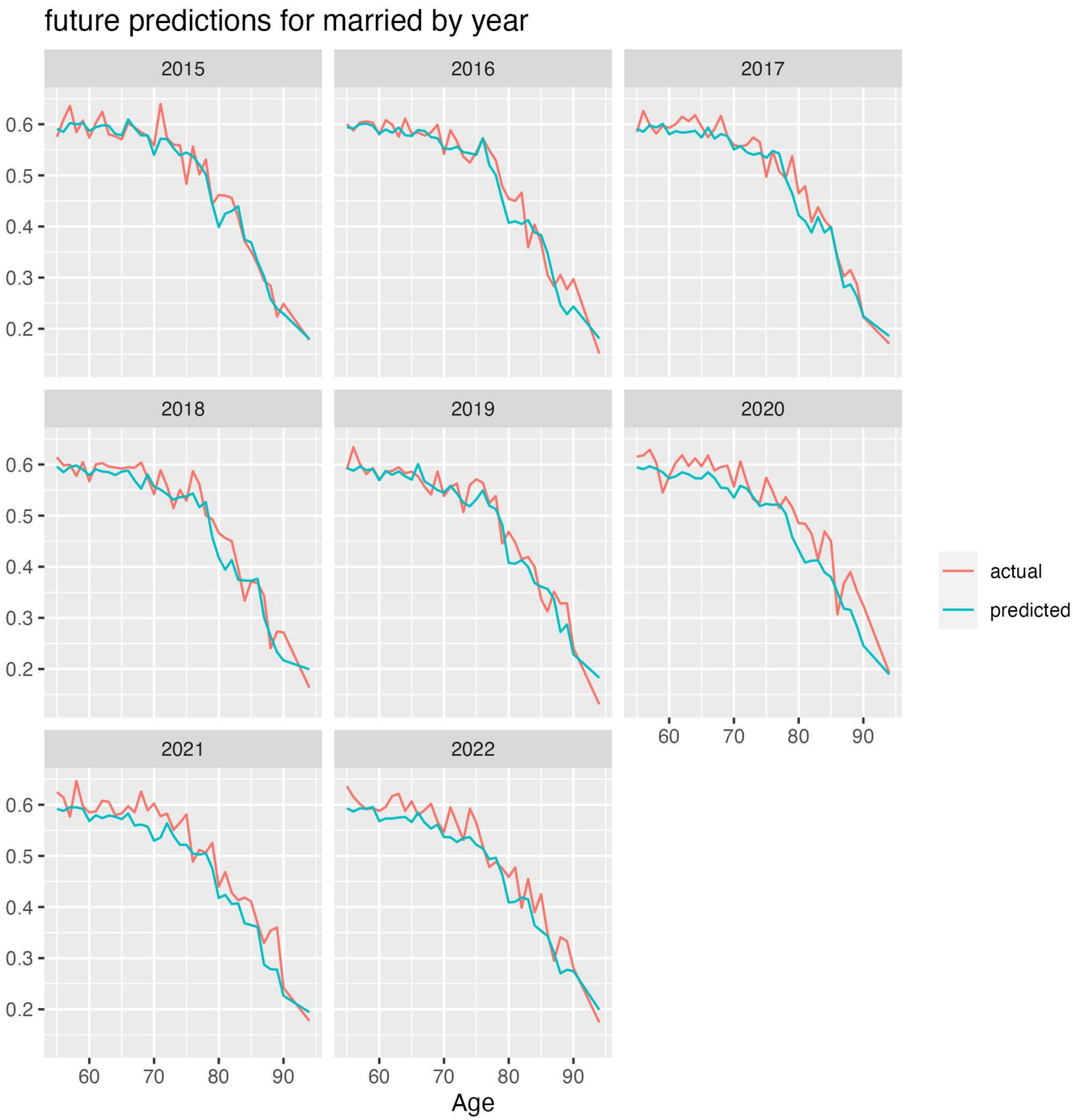
Predicted vs. actual from time split: home ownership by race, women only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B58

Predicted vs. actual from time split: married by year



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B59

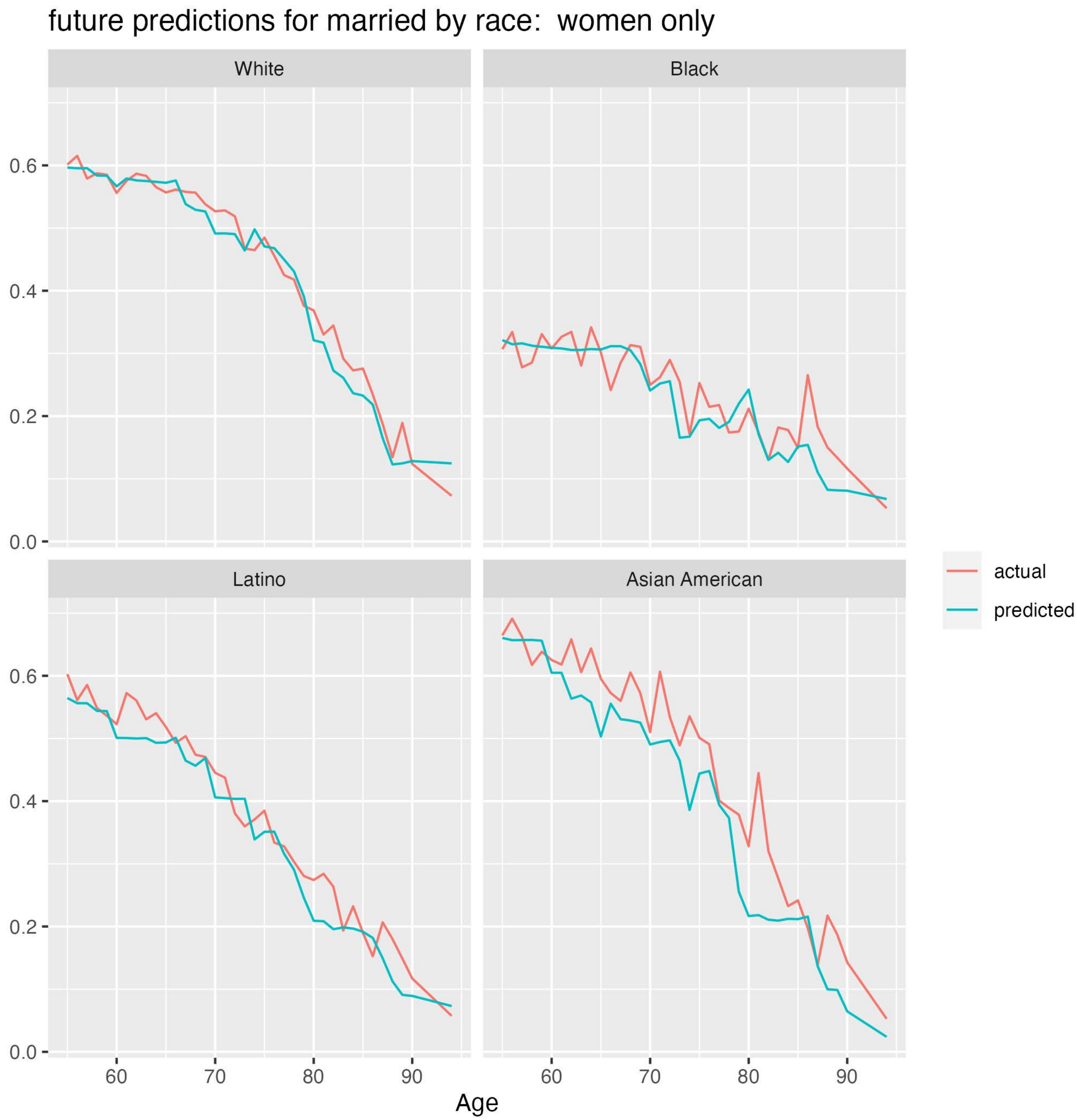
Predicted vs. actual from time split: married by race, men only

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B60

Predicted vs. actual from time split: married by race, women only

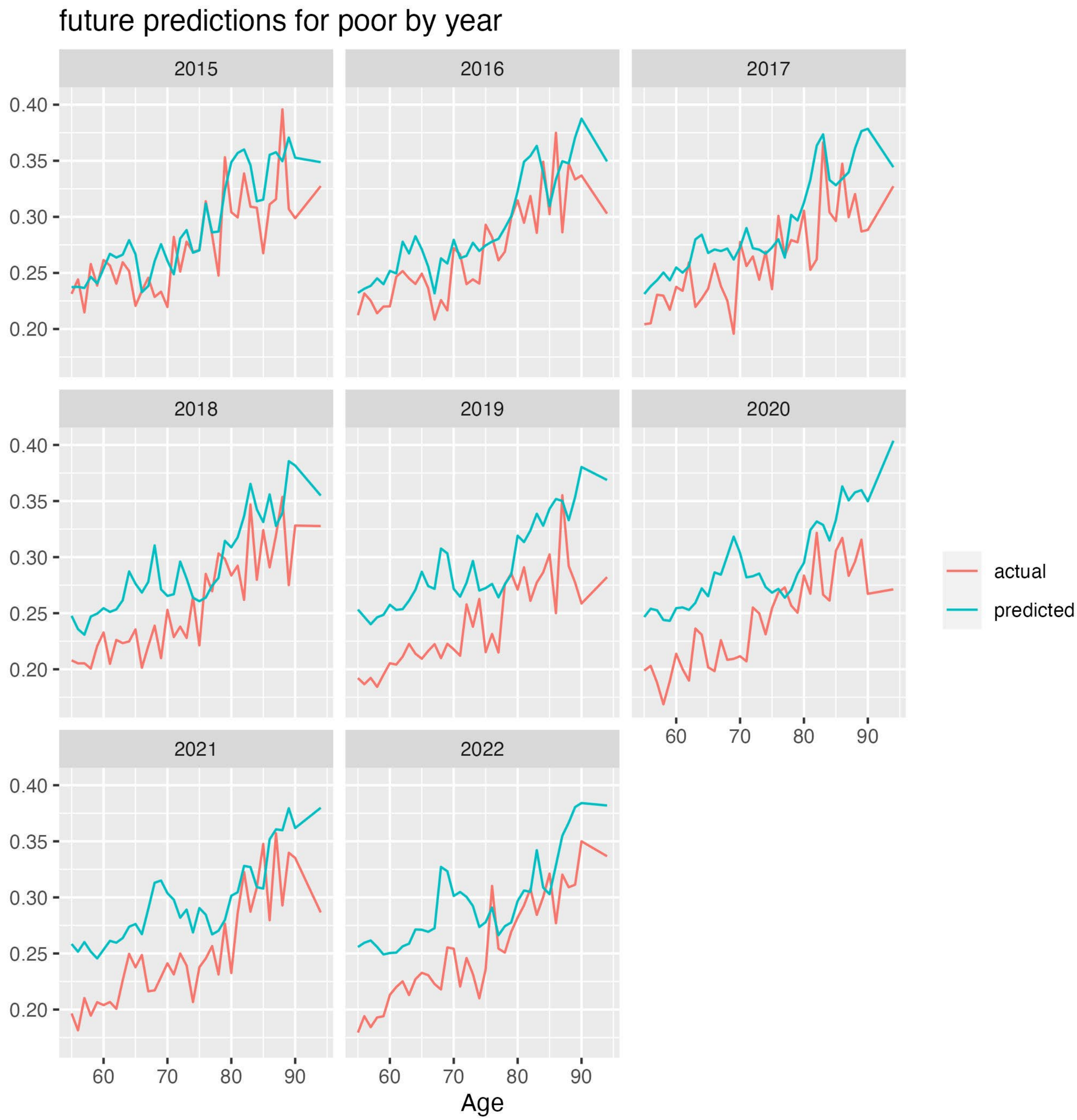


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B61

Predicted vs. actual from time split: poor (less than twice poverty level) by year

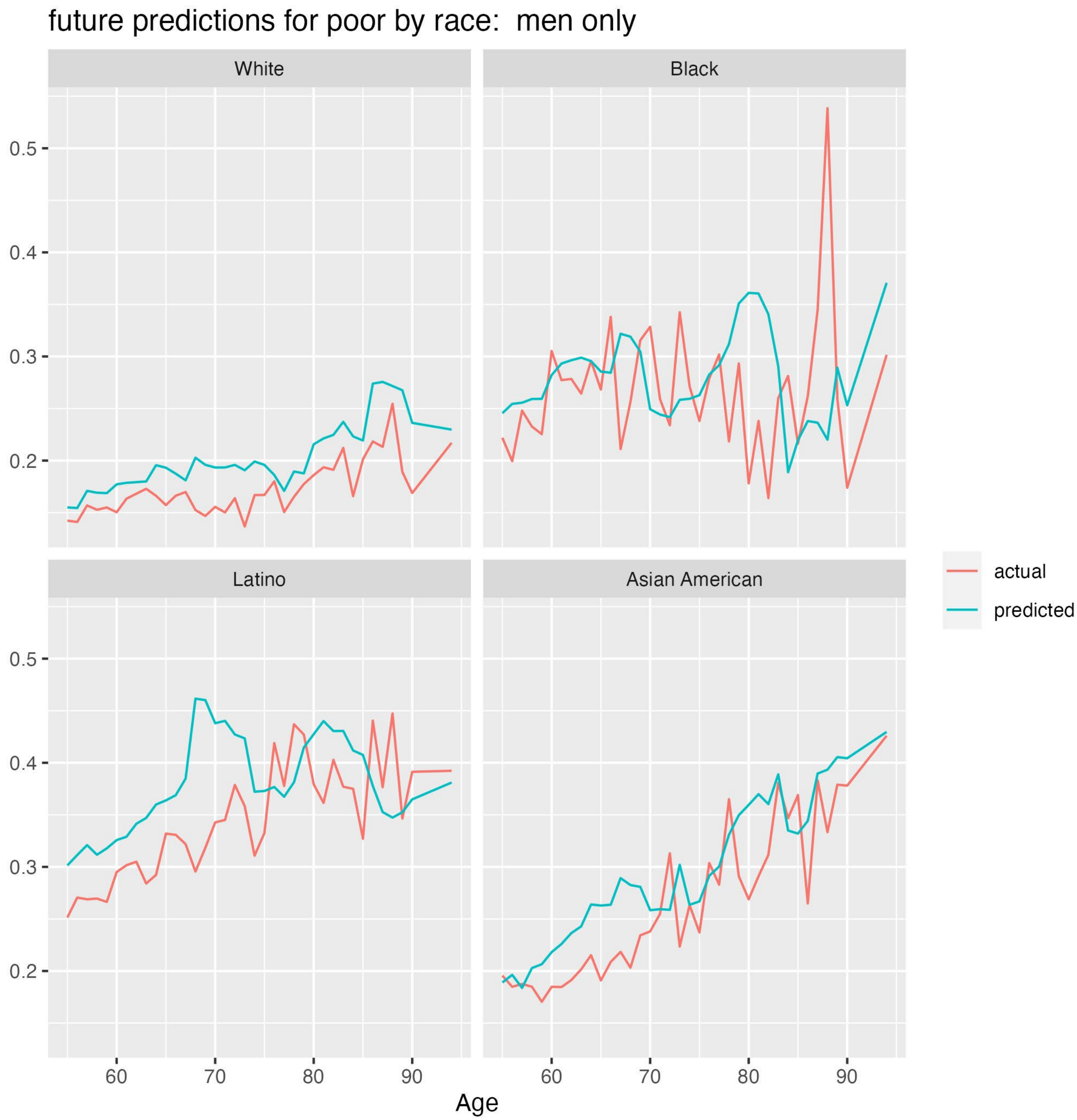


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B62

Predicted vs. actual from time split: poor (less than twice poverty level) by race, men only

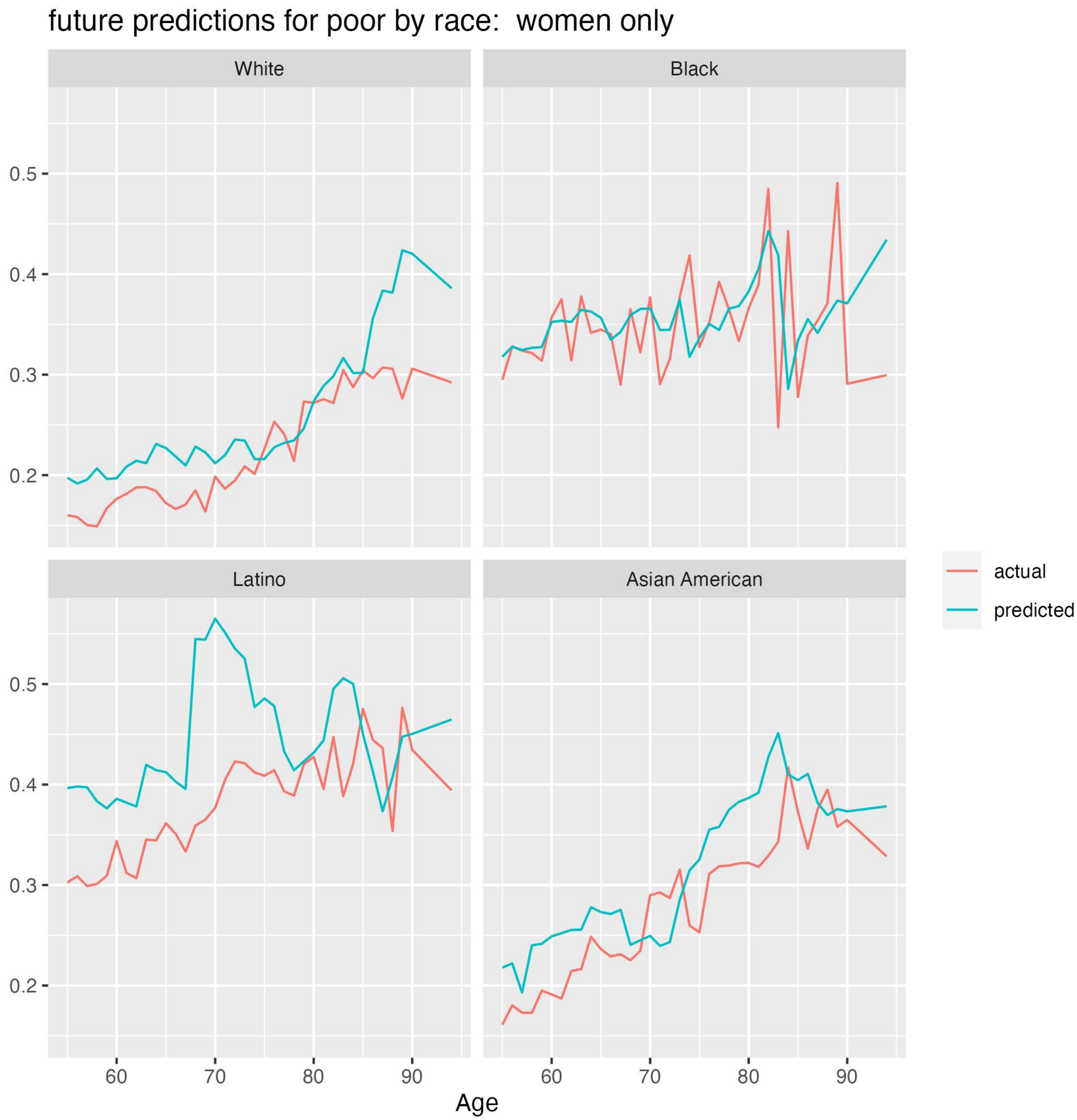


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B63

Predicted vs. actual from time split: poor (less than twice poverty level) by race, women only

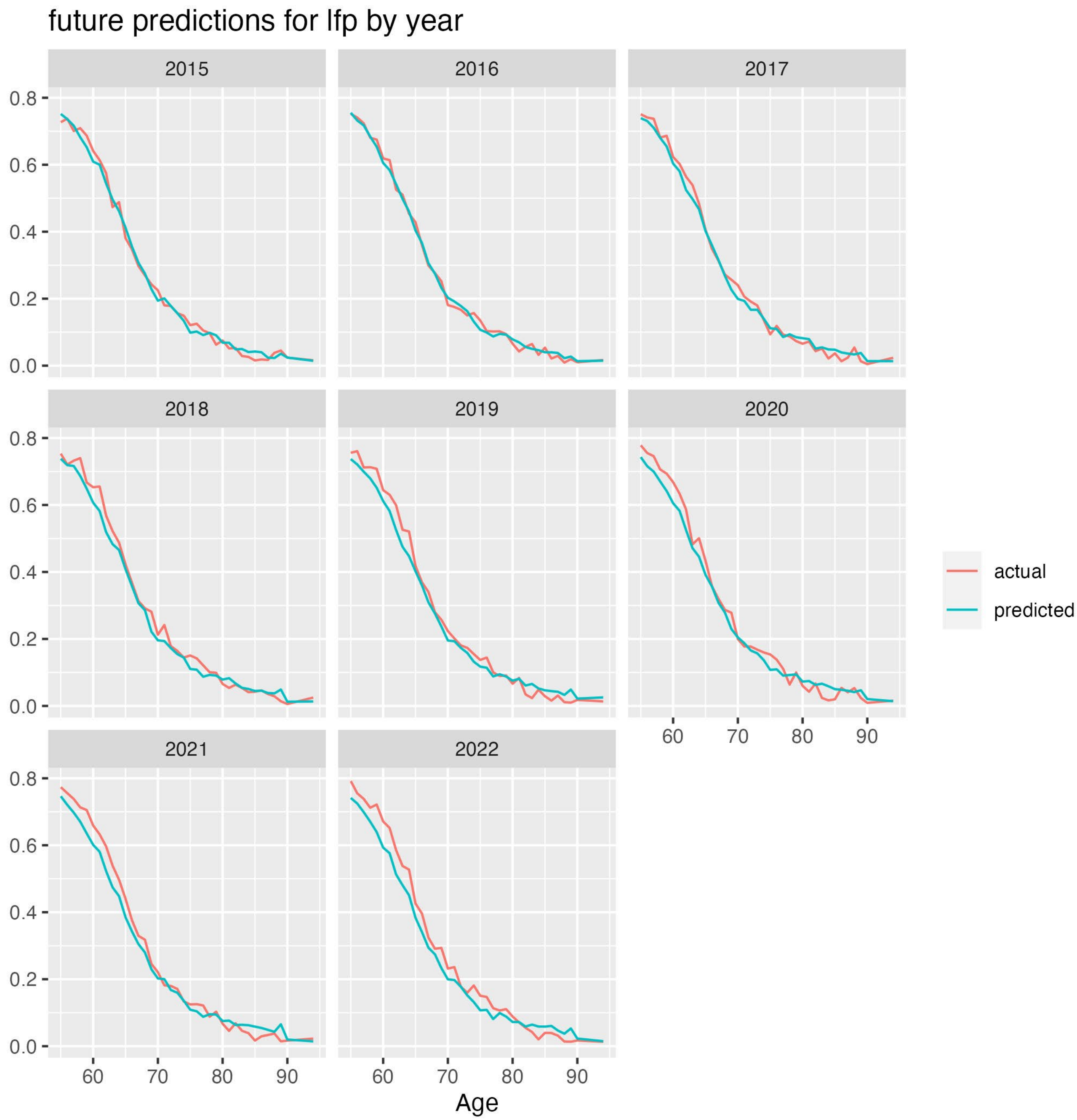


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B64

Predicted vs. actual from time split: labor force participation by year

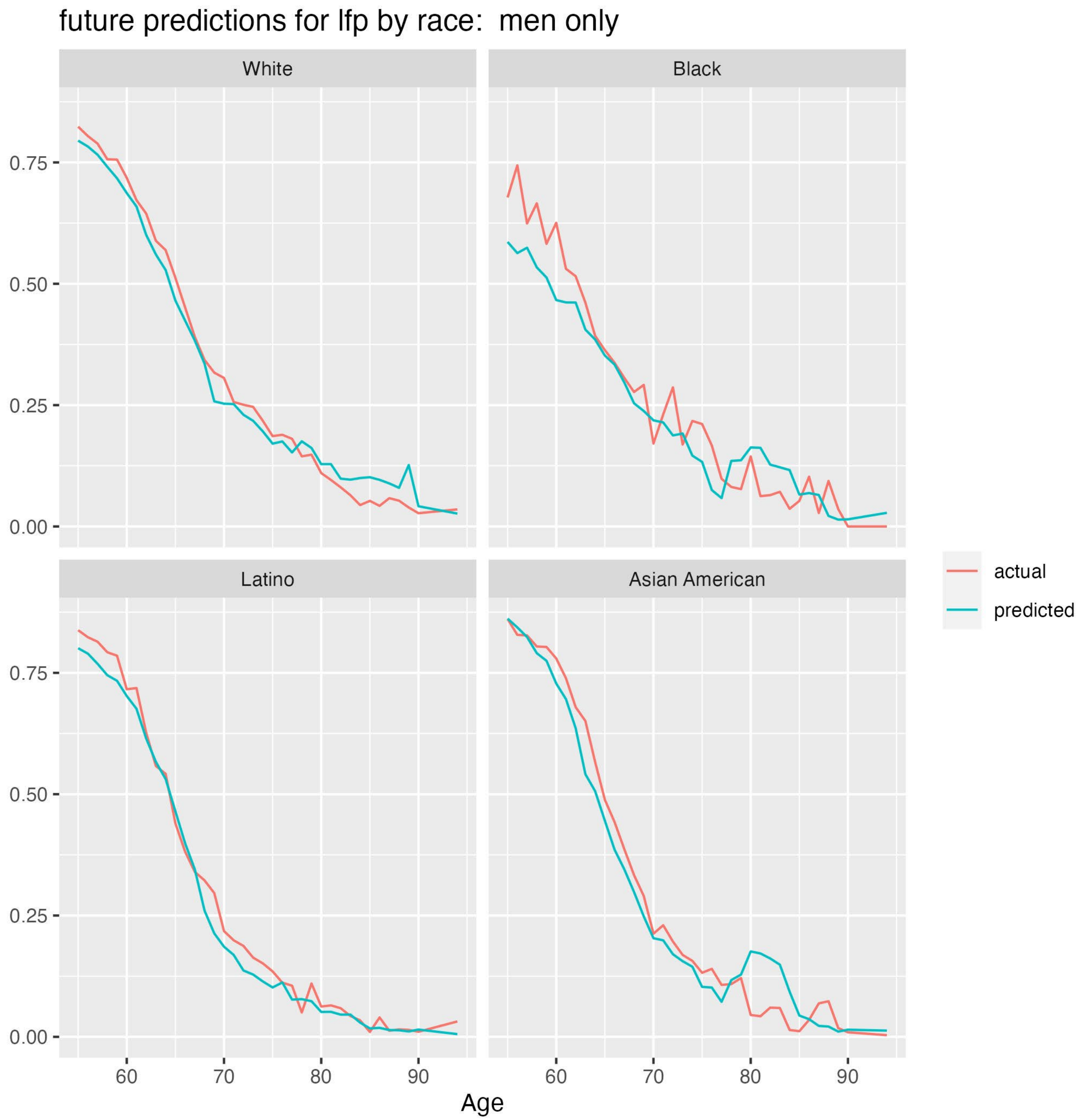


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B65

Predicted vs. actual from time split: labor force participation by race, men only

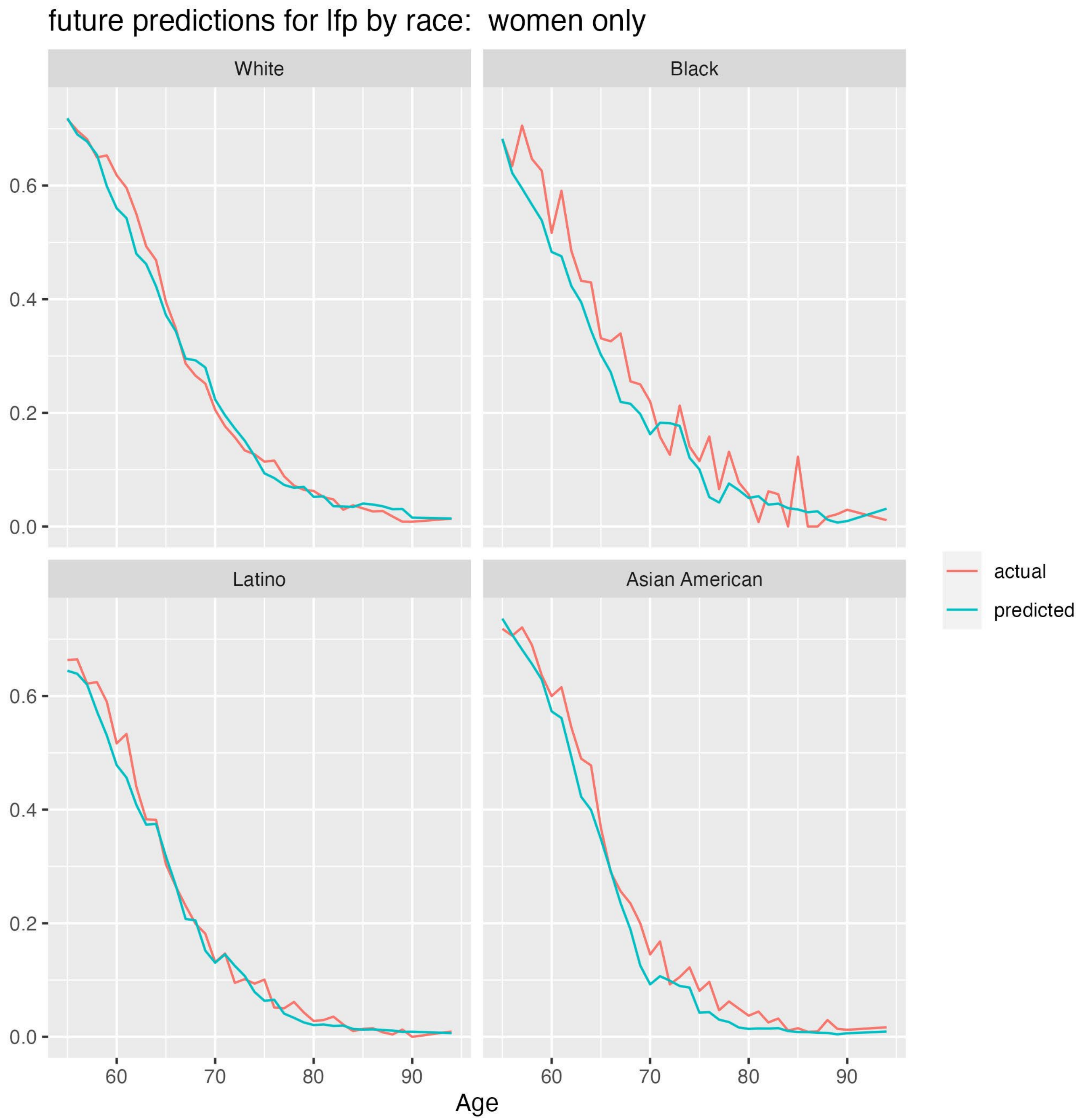


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B66

Predicted vs. actual from time split: labor force participation by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B67

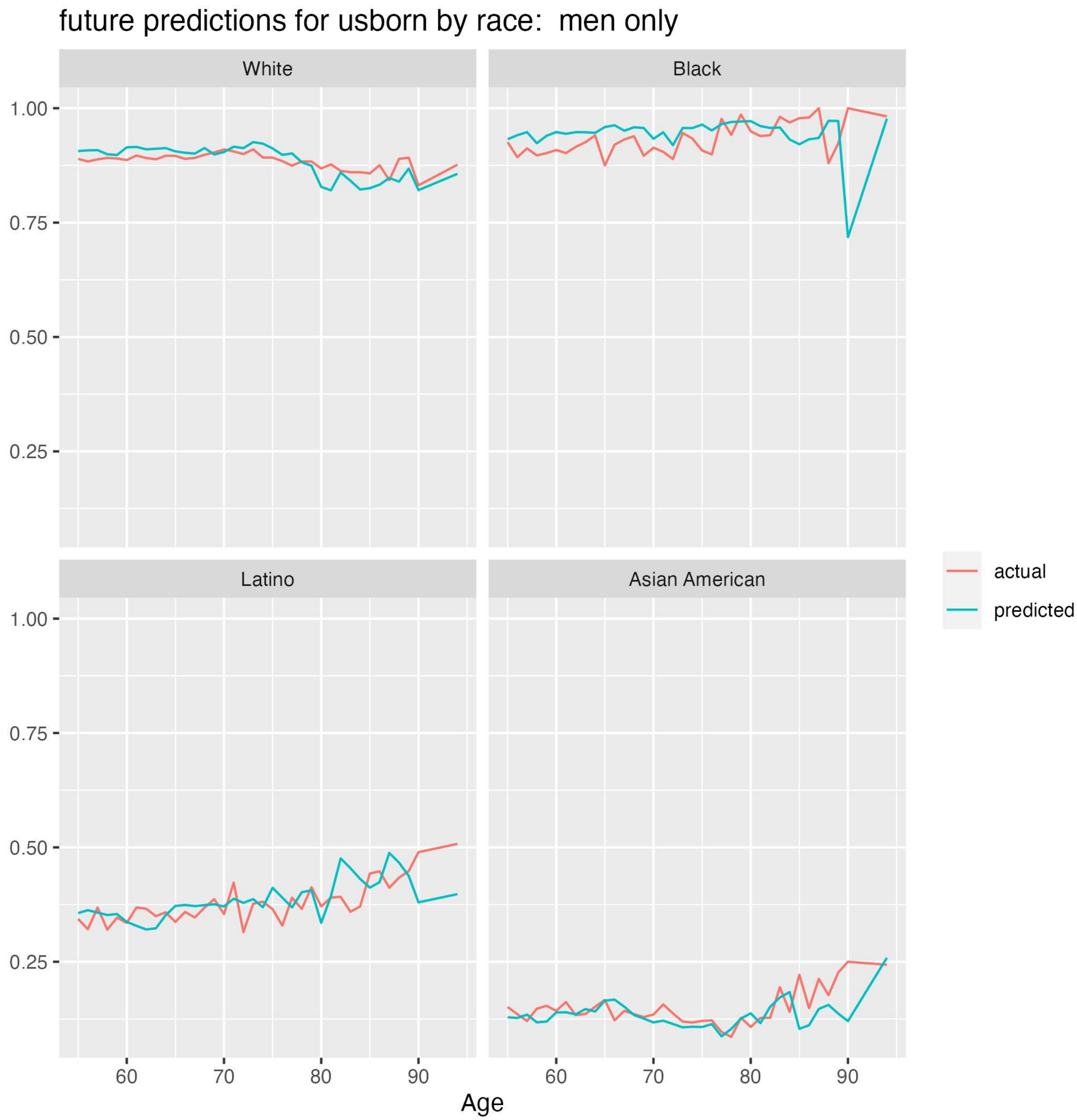
Predicted vs. actual from time split: U.S. born by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B68

Predicted vs. actual from time split: U.S. born by race, men only

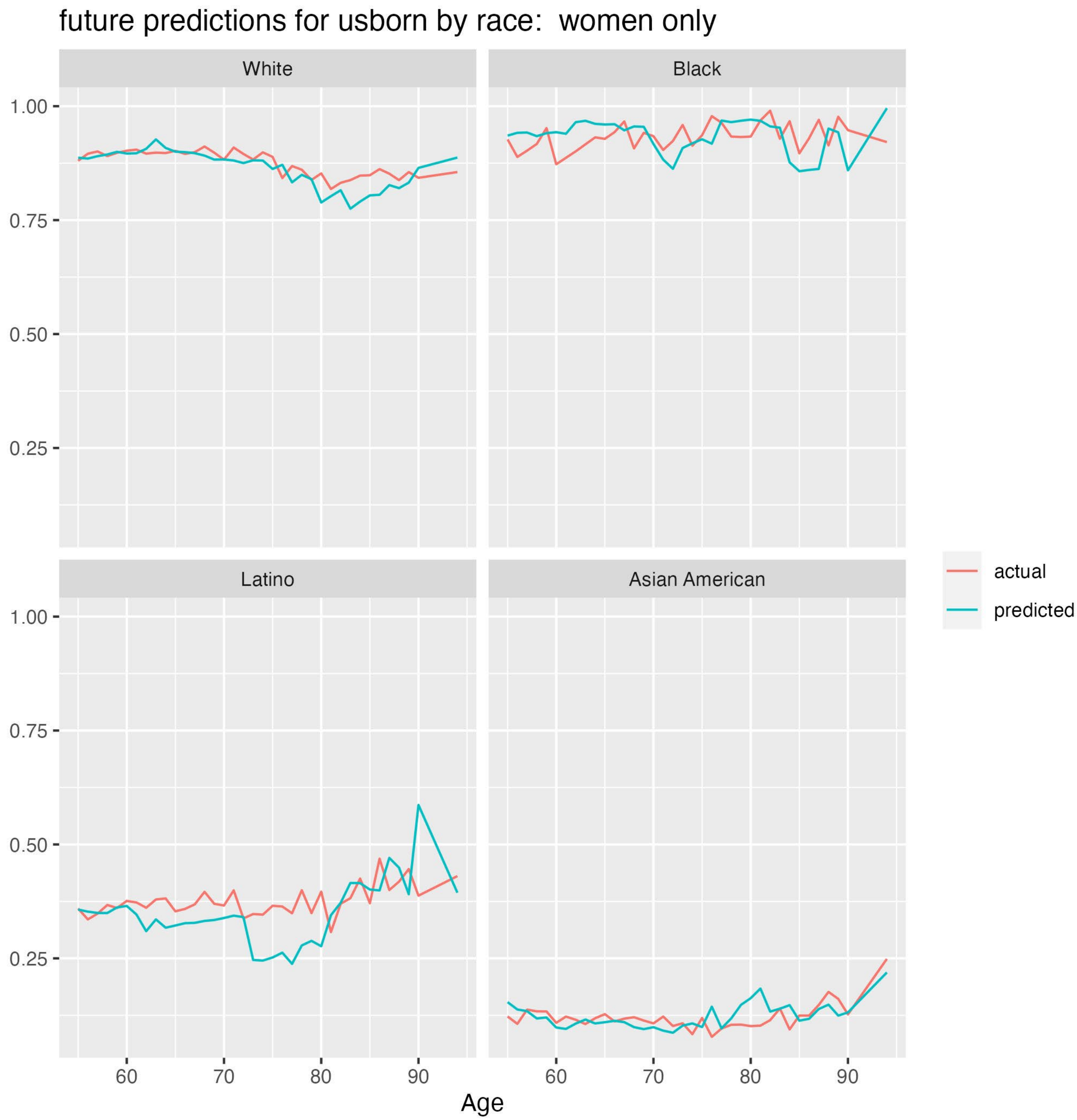


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B69

Predicted vs. actual from time split: U.S. born by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B70

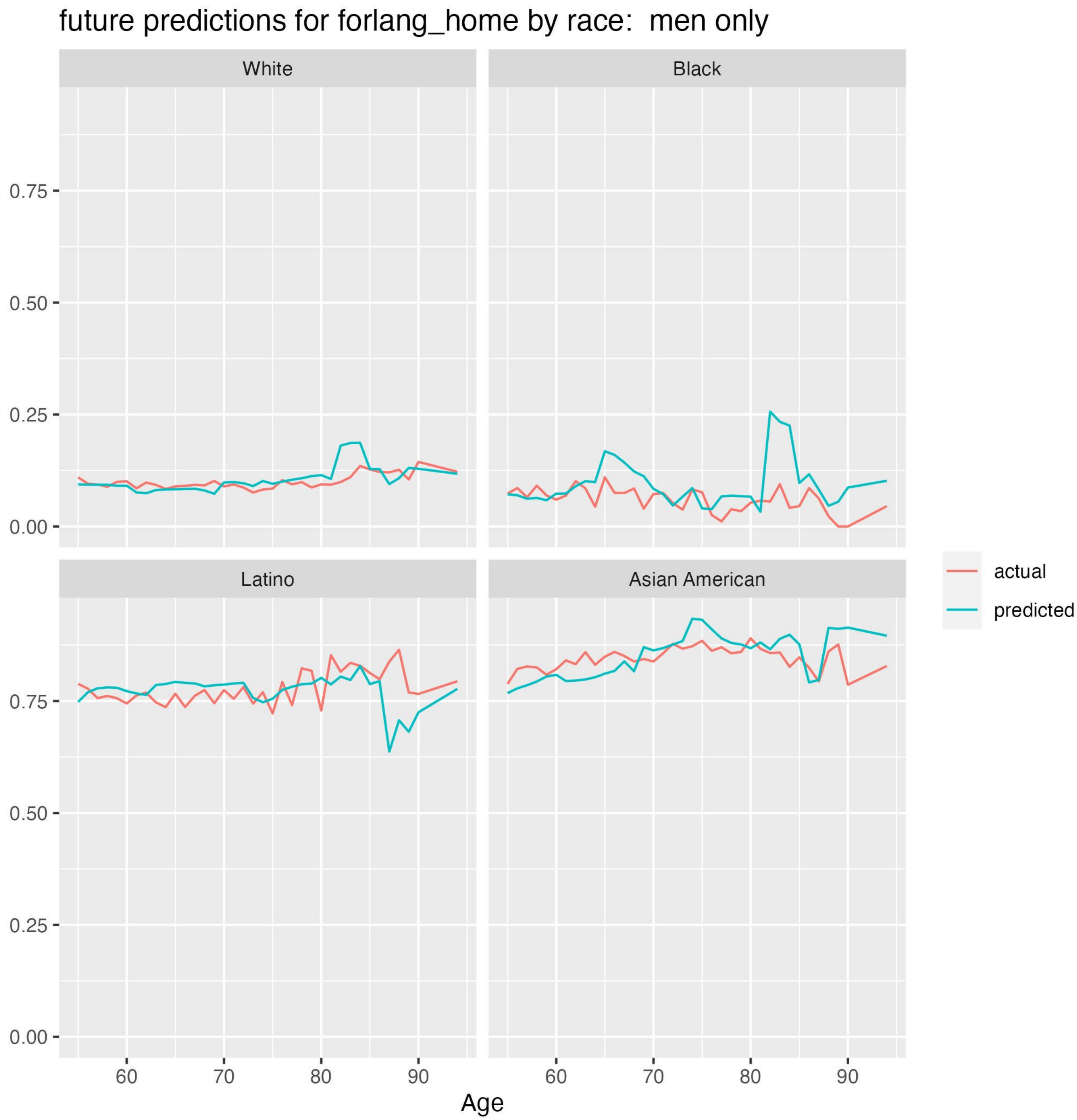
Predicted vs. actual from time split: speak foreign language at home by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B71

Predicted vs. actual from time split: speak foreign language at home by race, men only

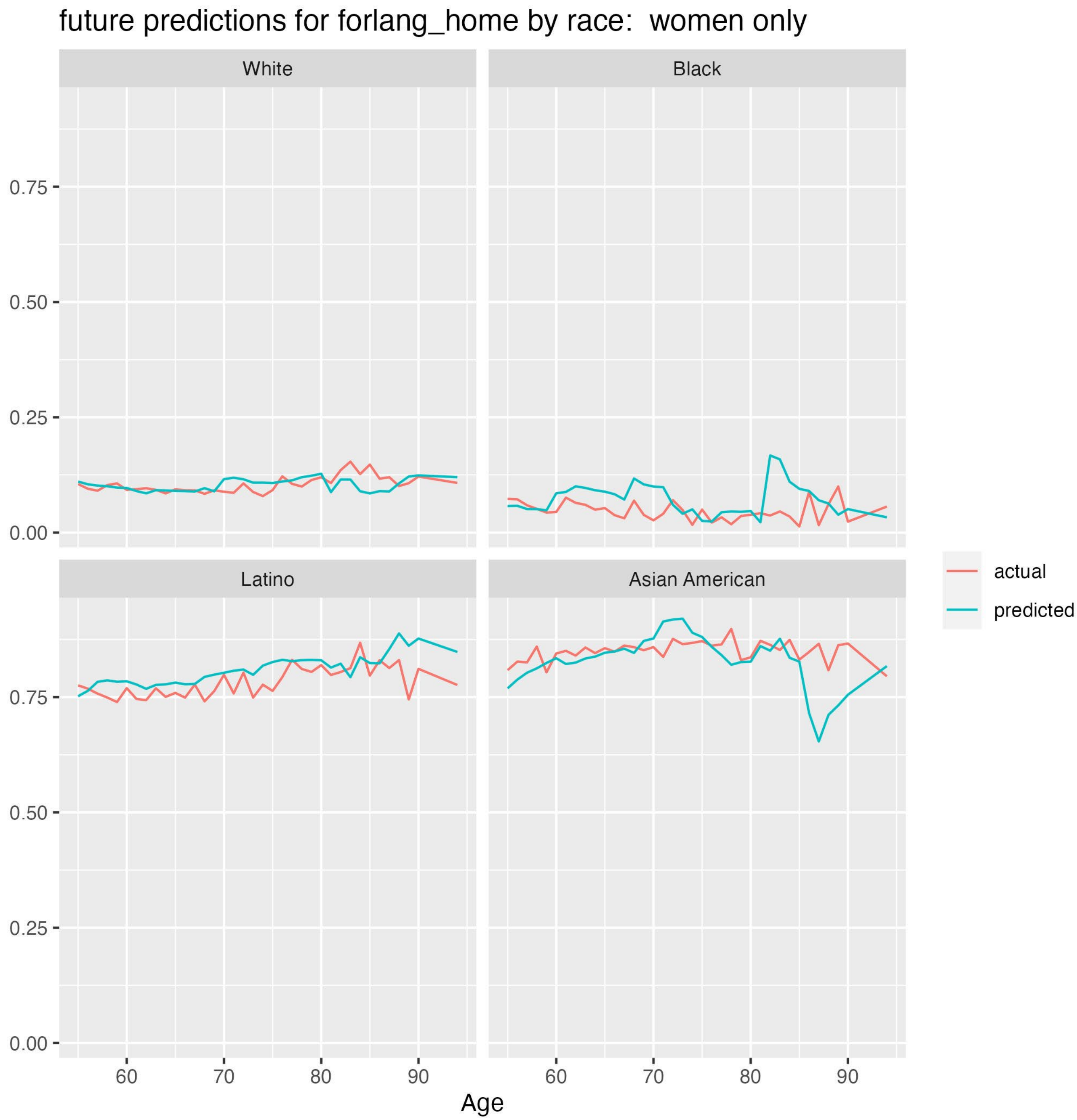


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B72

Predicted vs. actual from time split: speak foreign language at home by race, women only

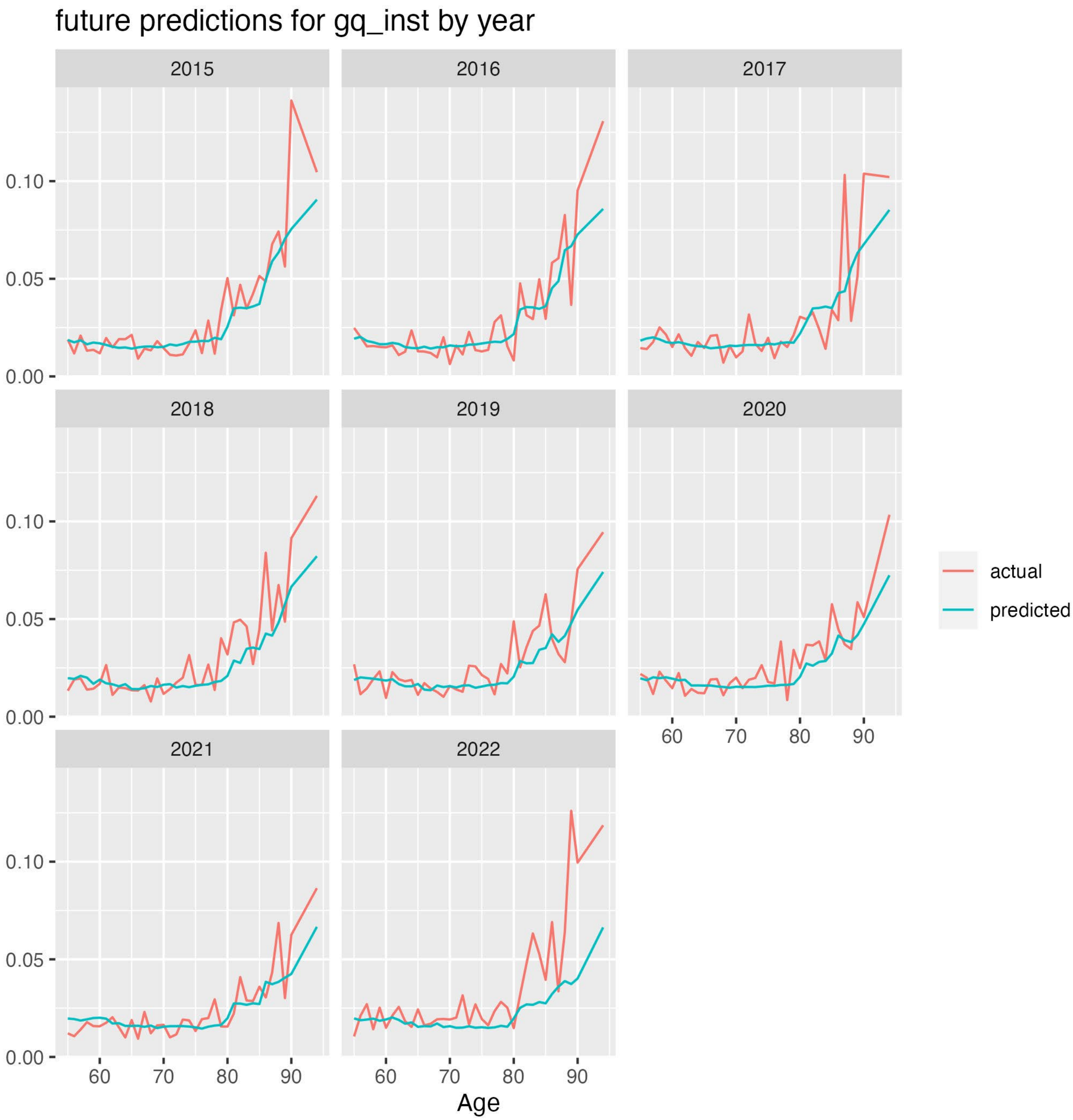


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B73

Predicted vs. actual from time split: live in institutional group quarters by year

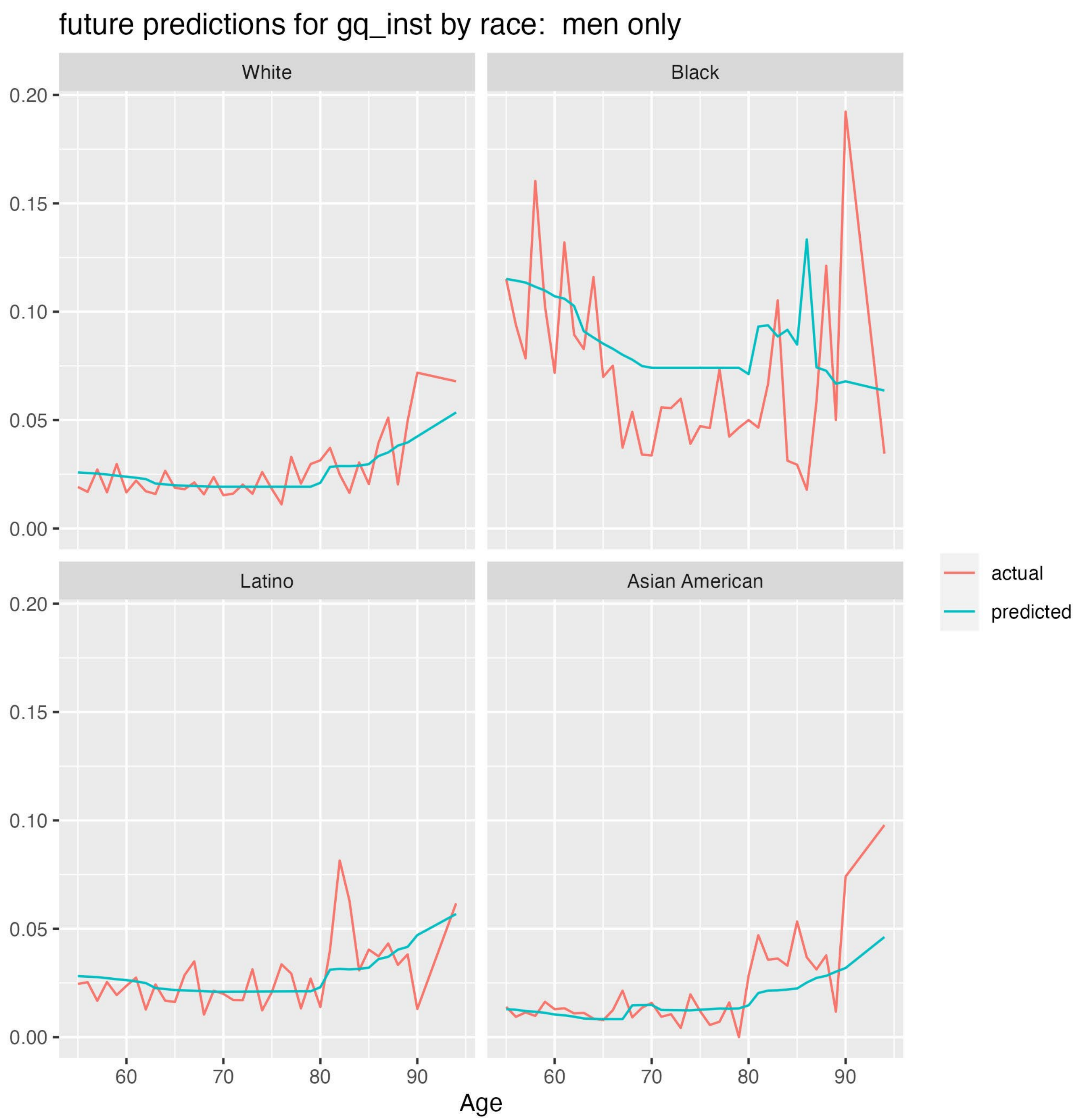


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B74

Predicted vs. actual from time split: live in institutional group quarters by race, men only

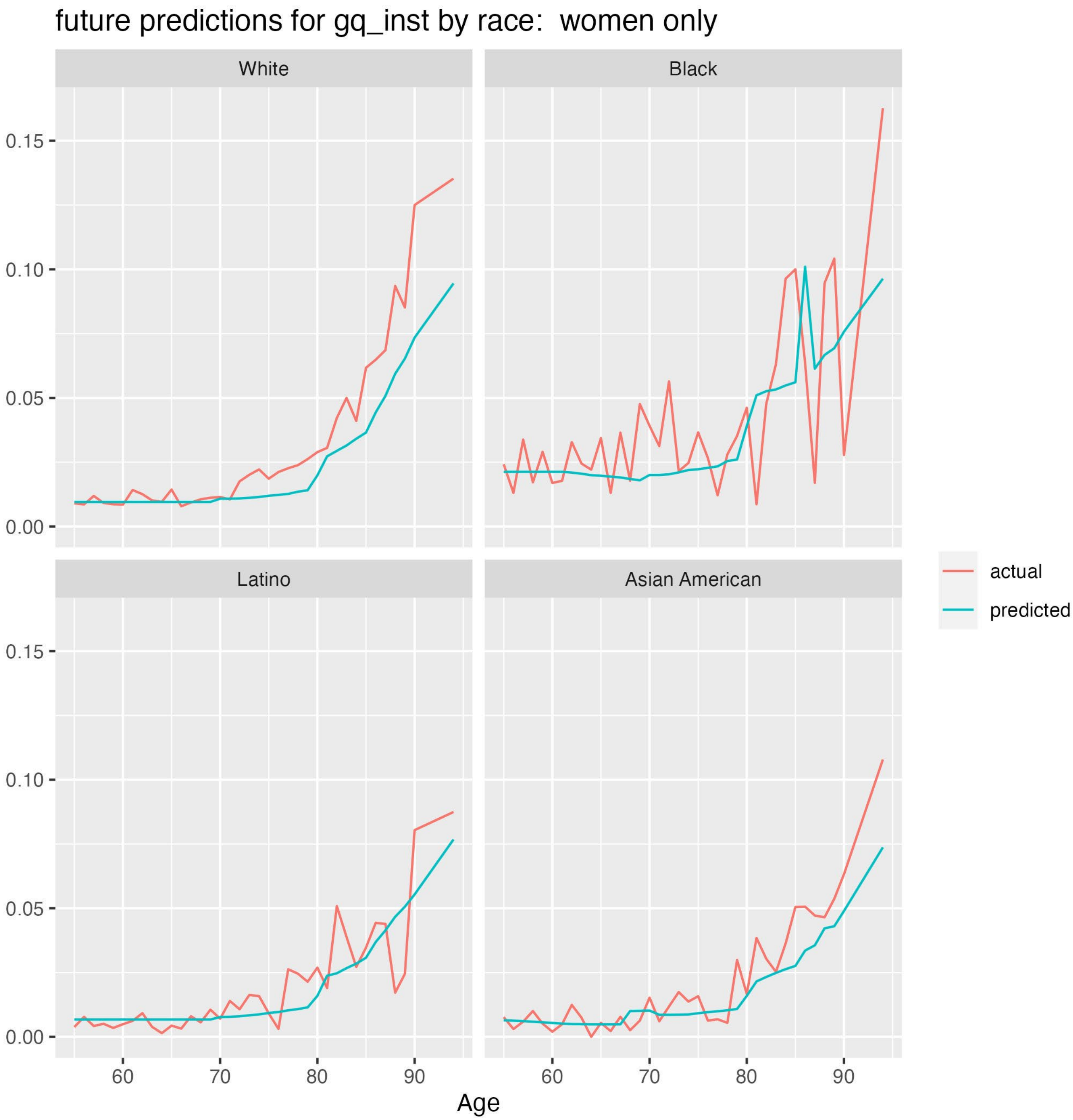


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B75

Predicted vs. actual from time split: live in institutional group quarters by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B76

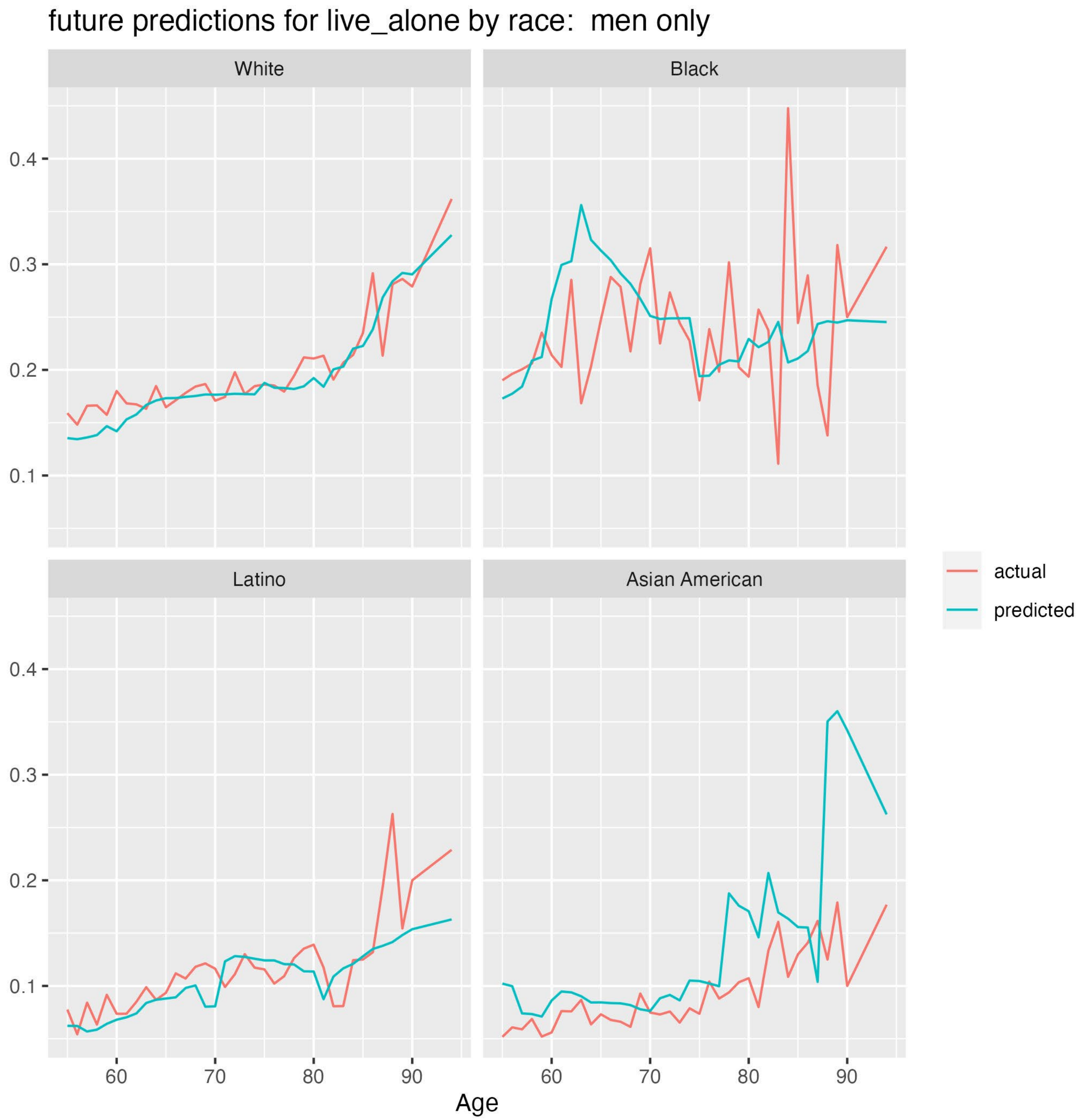
Predicted vs. actual from time split: live alone by year

SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B77

Predicted vs. actual from time split: live alone by race, men only

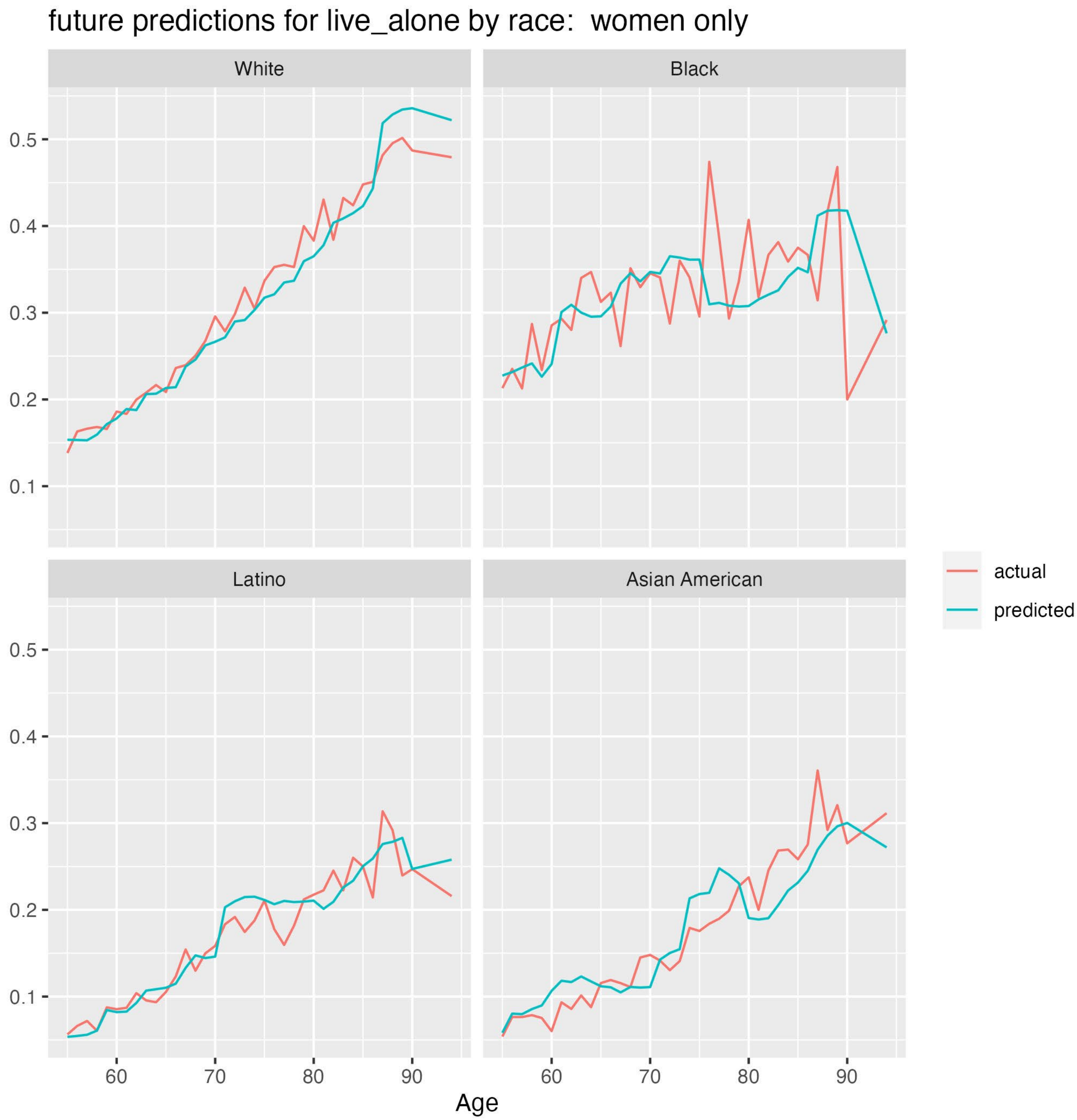


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B78

Predicted vs. actual from time split: live alone by race, women only

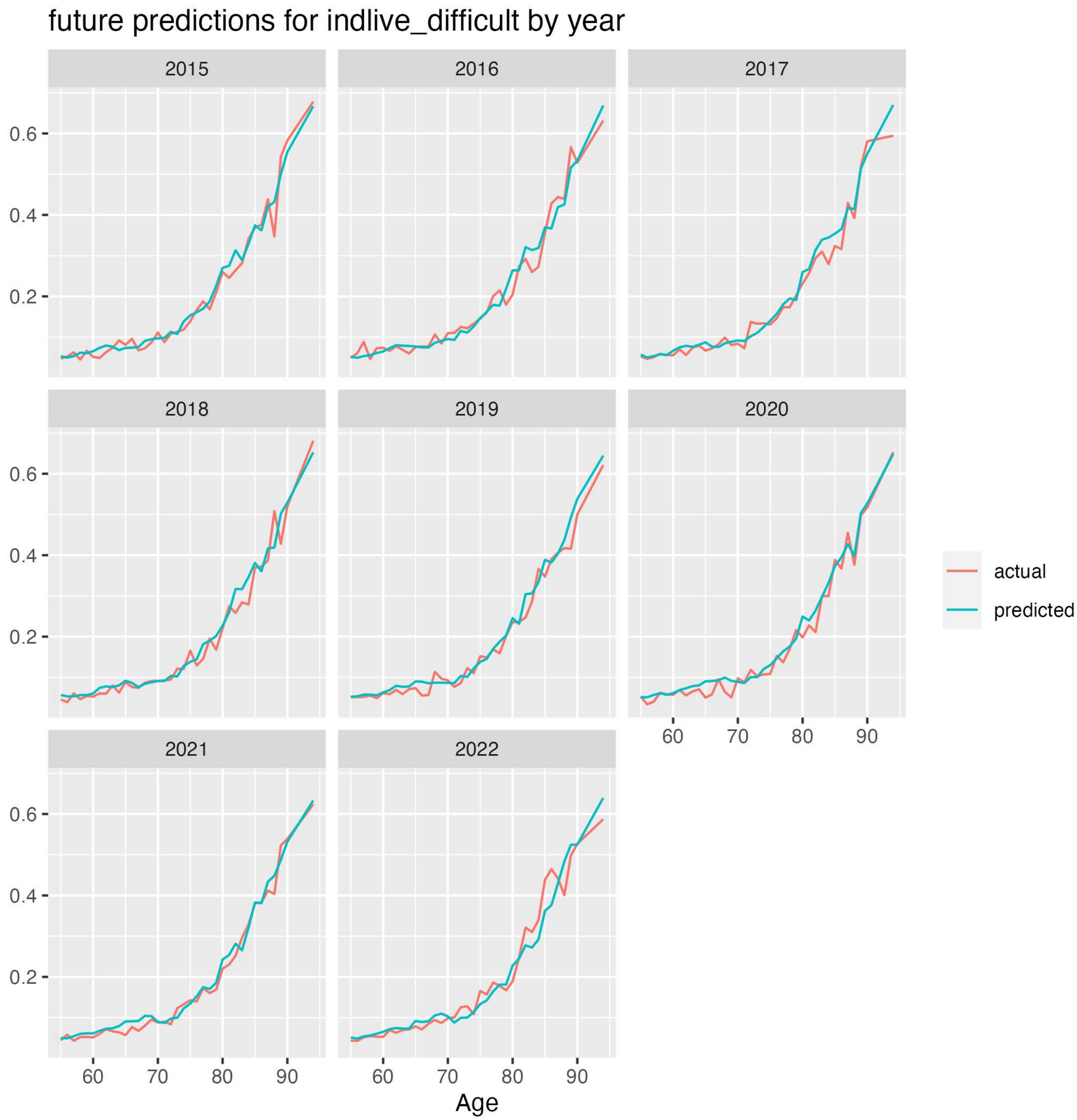


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B79

Predicted vs. actual from time split: difficulty living independently by year

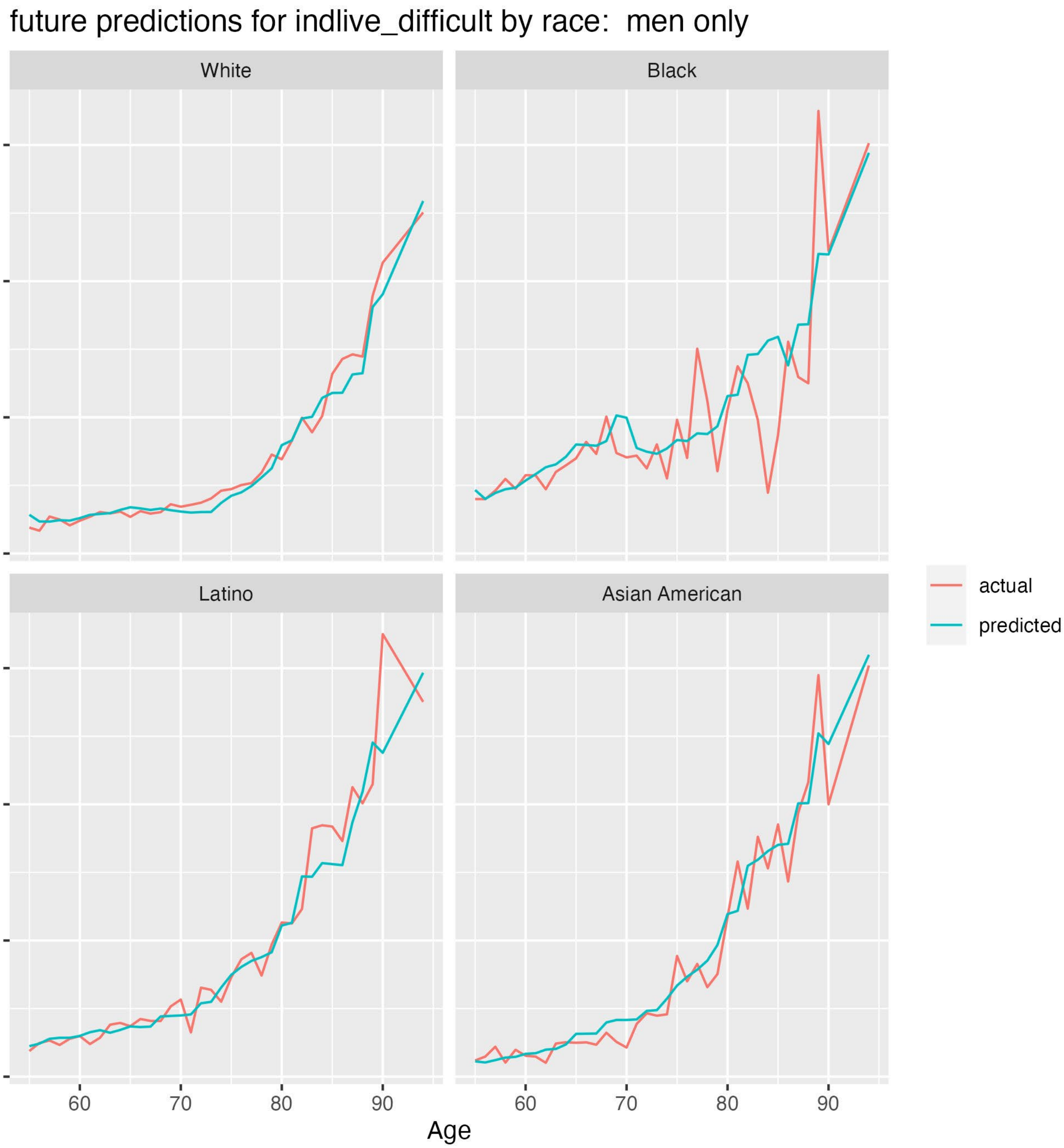


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B80

Predicted vs. actual from time split: difficulty living independently by race, men only

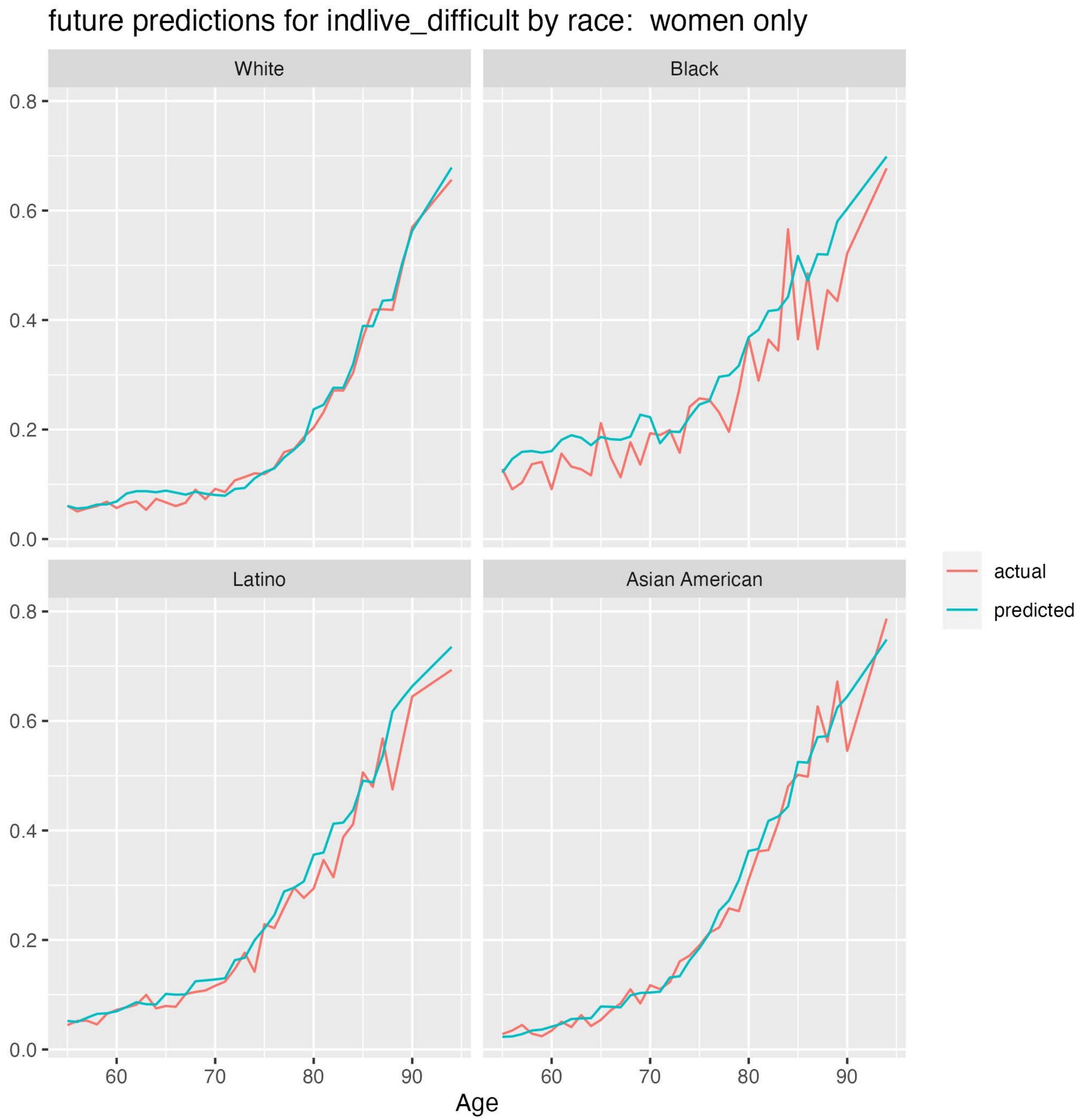


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B81

Predicted vs. actual from time split: difficulty living independently by race, women only

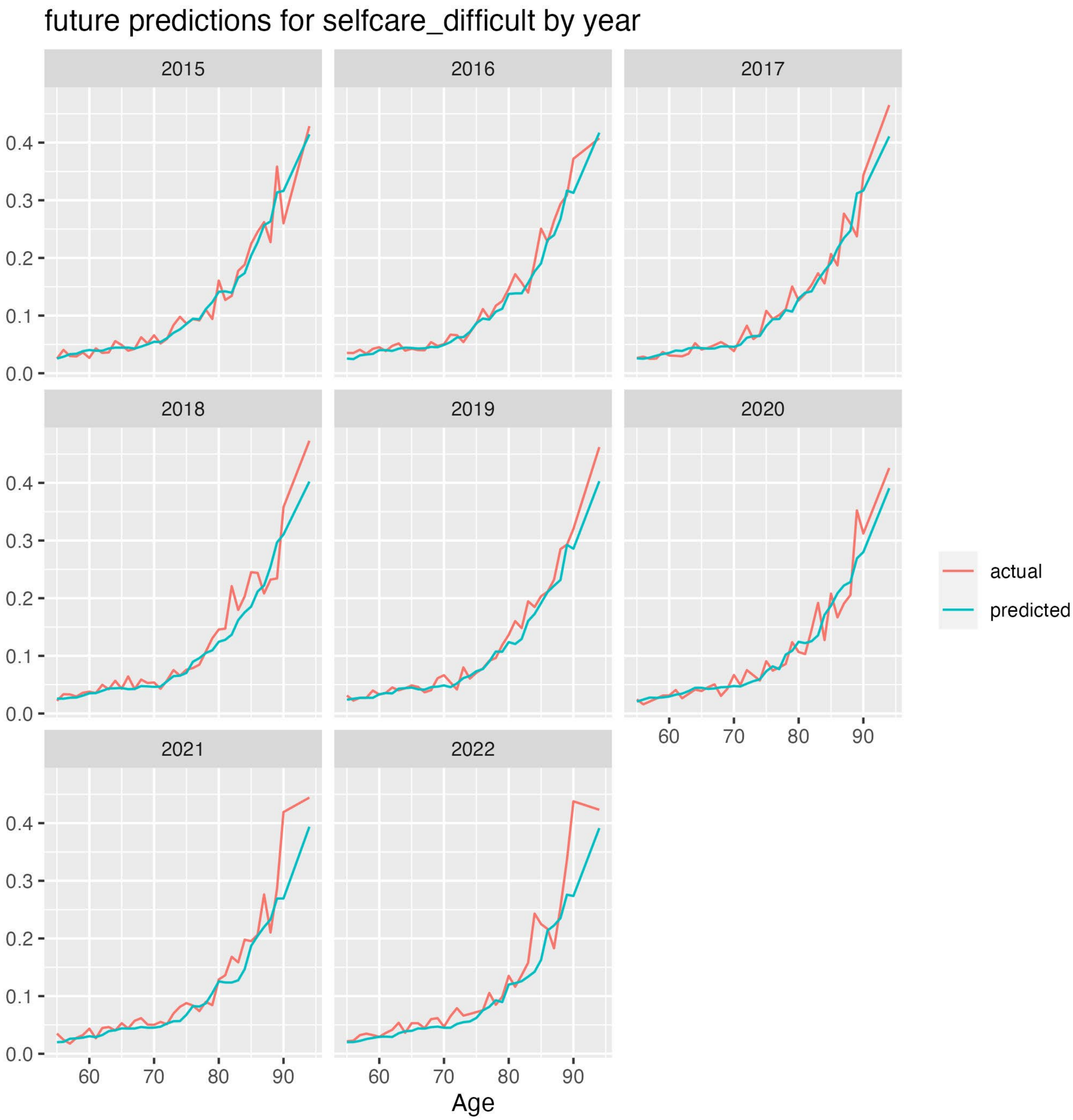


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B82

Predicted vs. actual from time split: difficulty with self care by year

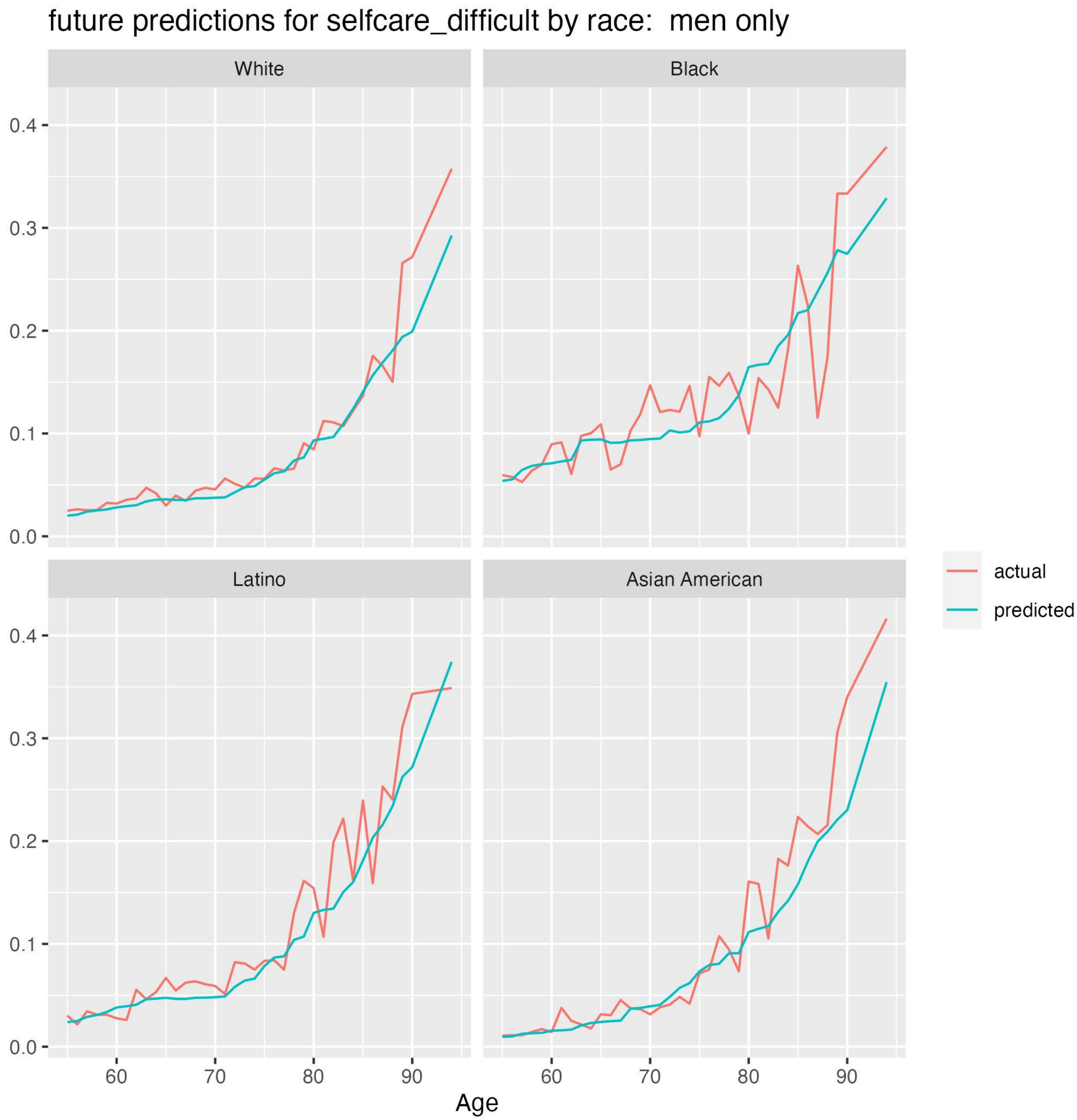


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B83

Predicted vs. actual from time split: difficulty with self care by race, men only

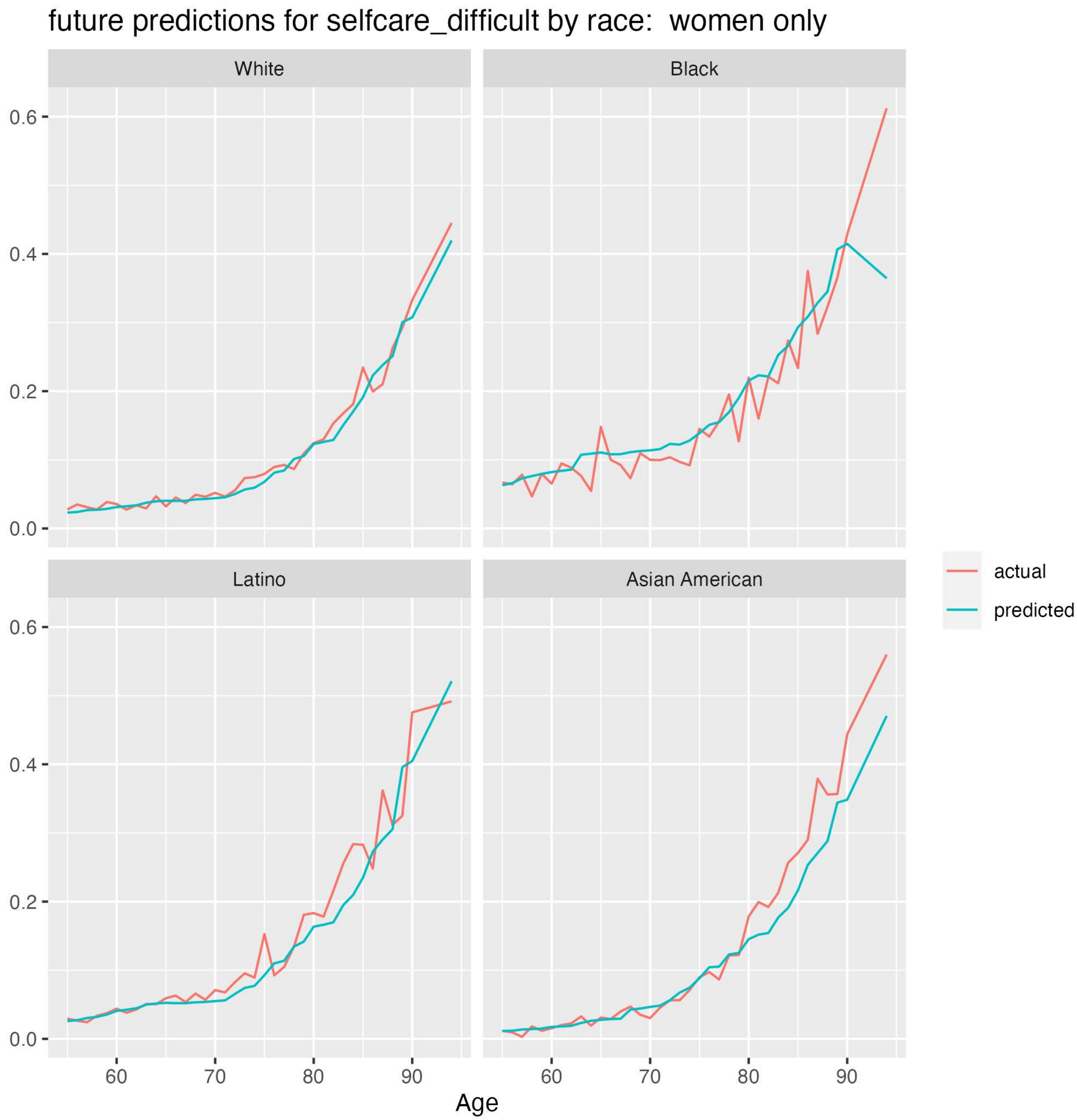


SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

FIGURE B84

Predicted vs. actual from time split: difficulty with self care by race, women only



SOURCE: American Community Survey from IPUMS USA https://usa.ipums.org/usa/cite.shtml.

NOTE: Model was trained on all the data from 2006 through 2014, and then tested on the data from 2015 through 2022.

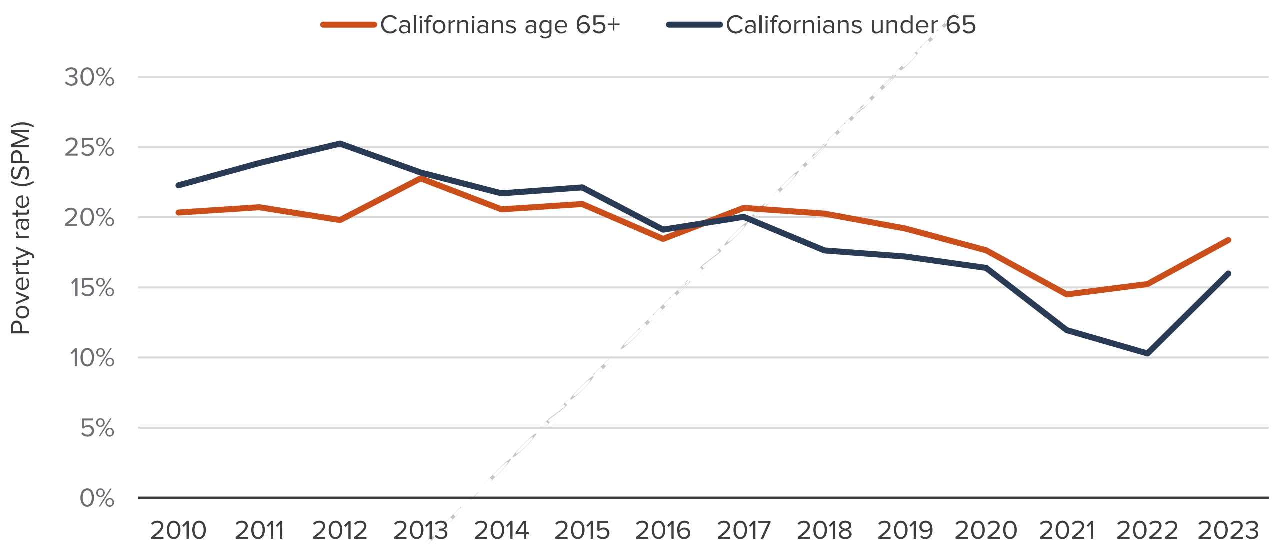
Appendix C. Additional information and figures

Measuring poverty

Supplemental poverty measures can provide a more holistic and accurate reflection of poverty because they take into account things like housing costs and safety net benefits – and importantly for older adults, medical expenses. The Current Population Survey includes supplemental poverty estimates, which are shown in Figure C1, for older Californians adults age 65 and older and those under 65. Prior to 2019, about one in five older California adults were considered in poverty according to the Census SPM. Poverty rates have since declined for both older adults and all Californians due to an improving economy, but even more so because of pandemic-related aid. However, the 2023 estimates suggest poverty may is growing once again.

FIGURE C1

Holistic measures of poverty suggest lower rates for older Californians in recent years



SOURCE: Current Population Survey

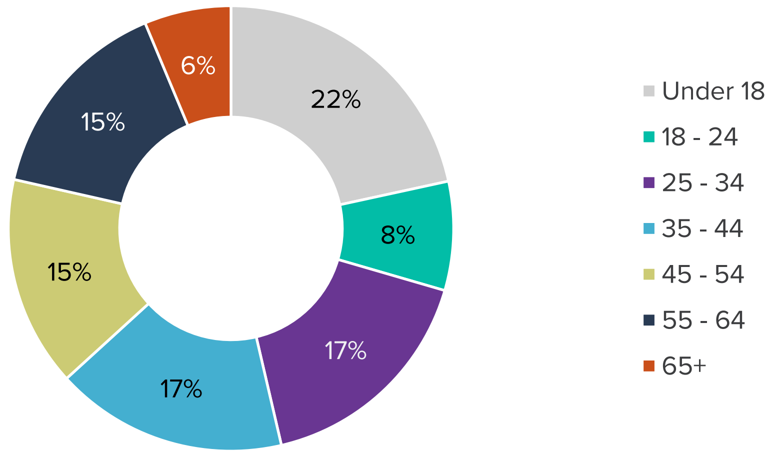
NOTE: Figure shows percent of Californians under 100% of poverty according to the SPM threshold.

Homelessness among older California adults

While older adults represent relatively small shares of people experiencing homelessness in California, the increase in homelessness among older adults in recent years has drawn increased policy attention and concern. According to data reported by local homelessness services agencies, **adults** over age 65 comprise a relatively small share – about 6% of Californians who received homelessness services in 2023 (Figure C2). Though a relatively small share, this translates into over 20,000 Californians over 65 experiencing homelessness. If we consider older adults as age 55 and older, they represent about one in five Californians experiencing homelessness. It may be that a higher share of older adults are experiencing homelessness but do not engage or receive any services from local homelessness agencies and are not included. Nonetheless, this is the best source of information we have on the age profile of Californians experiencing homelessness.

FIGURE C2

About 1 in 15 Californians experiencing homelessness are age 65 and older



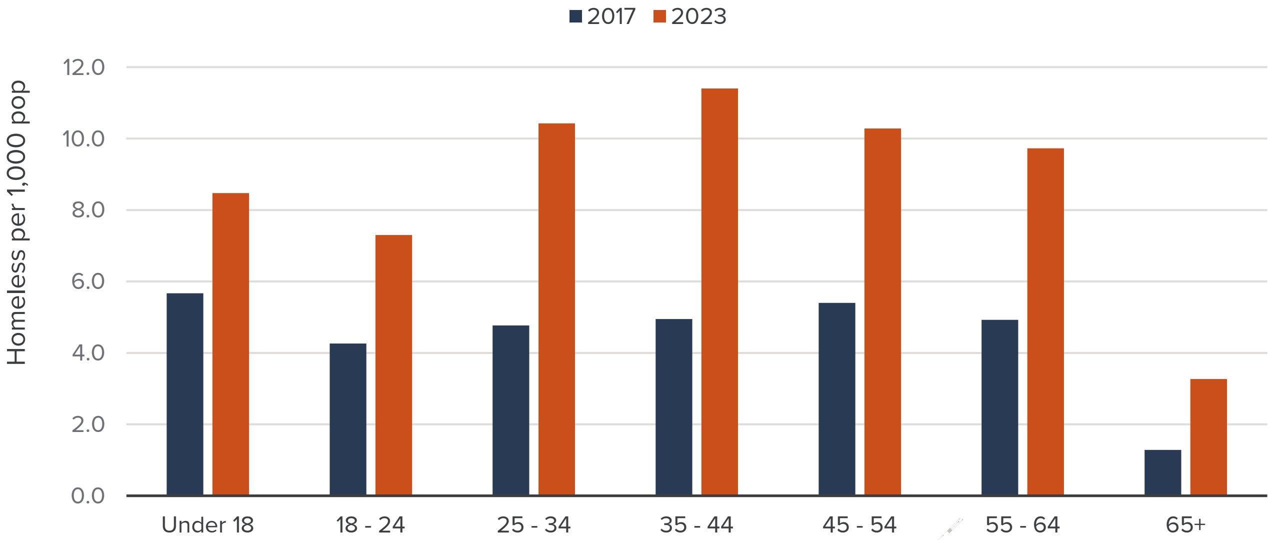
SOURCE: California Homeless Data Integration System, 2023

NOTE: Data reported by California's 44 Continuums of Care includes counts of people who received homelessness services in 2023.

About 3 out of 1000 Californians 65 and older experienced were experiencing homelessness in 2023 (Figure C3). This is much lower than per capita rates of homelessness than any other age group. However, the rate for adults 55 to 64 is more than double that at about 10 out of 1000 experiencing homelessness. It is also the case that adults over 65 have experienced the largest growth since 2017, which is the first year we have statewide homelessness counts by age available. In 2017, only about 1 in 1000 adults 65 and older received homelessness services in California compared to about 3 in 1000 in 2023.

FIGURE C3

Adults over 65 have the lowest rates of homelessness per capita, but have more than doubled in 6 years



SOURCE: California Homeless Data Integration System; California Department of Finance.

NOTE: Data reported by California's 44 Continuums of Care includes counts of people who received homelessness services in 2017 and 2023. Total population counts by age come from DOF population estimates for 2017 and 2023.

TABLE C1

Characteristics of older adults in California, 2022

	Age 55 - 64	Age 65 and older	Age 65 - 74	Age 75 - 84	Age 85 and older
Median income	$108,100	$79,300	$86,000	$72,800	$59,400
Own home, no mortgage	21.6%	35.6%	32.0%	39.5%	42.9%
Own home, with mortgage	47.7%	37.9%	41.9%	35.4%	24.5%
Renter	28.7%	23.6%	24.1%	22.5%	24.3%
Labor force participation	66.5%	19.4%	28.0%	9.6%	2.3%
Low-income (under 200% FPL)	22.2%	28.1%	25.1%	29.5%	39.3%
Living alone	14.3%	24.6%	20.4%	27.2%	38.7%
Living with spouse	27.8%	23.3%	26.0%	22.2%	12.5%
Living with family members	16.0%	16.3%	15.1%	16.5%	22.0%
Any disability	15.8%	34.4%	23.7%	40.9%	71.2%
Self-care difficulties	3.4%	10.4%	5.7%	11.6%	31.3%
Independent living difficulties	5.9%	17.4%	9.3%	20.5%	49.9%
Total population	4,682,656	6,165,865	3,580,994	1,870,932	713,939

SOURCE: American Community Survey, 1-year PUMS, 2022

PUBLIC POLICY
INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.

Public Policy Institute of California
500 Washington Street, Suite 600
San Francisco, CA 94111
T: 415.291.4400
F: 415.291.4401
PPIC.ORG

PPIC Sacramento Center
Senator Office Building
1121 L Street, Suite 801
Sacramento, CA 95814
T: 916.440.1120
F: 916.440.1121