Black–white differences in intergenerational economic mobility in the United States

Bhashkar Mazumder

Introduction and summary

The large and persistent gap in economic status between blacks and whites in the United States has been a topic of considerable interest among social scientists and policymakers for many decades. The historical legacy of slavery and segregation raises the question of how long black Americans are likely to remain a disadvantaged minority. Despite the enormous literature on black–white inequality and its historical trends, few studies have directly measured black–white differences in rates of intergenerational mobility, that is, the ability of families to improve their position in the income distribution from one generation to the next. Estimates of rates of intergenerational mobility by race can provide insight on whether racial differences in the United States are likely to be eliminated and, if so, how long it might take. Furthermore, they might also help inform policymakers as to whether there are lingering racial differences in equality of opportunity and, if so, what the underlying sources for these differences are.

More generally, the relatively low rate of intergenerational mobility in the United States compared with other industrialized countries has been a growing concern to policymakers across the political spectrum.[1] Understanding the sources of racial differences in intergenerational mobility might also shed light on the mechanisms behind the relatively high degree of intergenerational persistence of inequality in the United States.

In this study, I attempt to advance our understanding along several dimensions. First, I use two data sets containing larger intergenerational samples than have been used in the previous literature. One of the data sets matches individuals in the U.S. Census Bureau's *Survey of Income and Program Participation* (SIPP) to administrative earnings records from the Social Security Administration (SSA). This matched data set provides many more years of data on parents' earnings than most surveys and is likely to be less prone to measurement error, since it is derived from tax records. In addition, the SIPP contains data on

Bhashkar Mazumder is a senior economist and research advisor and the director of the Chicago Census Research Data Center at the Federal Reserve Bank of Chicago.
The author thanks Katherine Meckel, Nathan Chan, Benjamin Jakes, Weiting Zhang, and especially Jon Davis for excellent research assistance. He thanks Nathan Grawe, Gregory Clark, Derek Neal, Jason Faberman, and Dick Porter for their helpful suggestions and seminar participants at the University of California, Davis 2010 Conference on Social Mobility, the Federal Reserve Bank of Chicago, the 2011 ASSA (Allied Social Science Associations) Meeting, and the 2011 World Bank/IAE Conference on Socio-Economic Mobility and the Middle Class in Latin America for helpful comments. Any opinions and conclusions expressed herein do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Economic Perspectives is published by the Economic Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

ISSN 0164-0682

key characteristics of the parents, such as wealth levels and marital history. The other data source I use is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth* (NLSY). In addition to containing a rich array of information as children transition from adolescence to adulthood, such as test scores and personality traits, the NLSY also measures total family income in both generations, giving it an advantage over the SIPP. Using a measure of economic status that includes the income of the spouse avoids selecting only women who participate in the labor market.

Second, I use two types of measures of intergenerational mobility. The first is a set of transition probabilities of relative income status across generations. An example is the probability of moving out of the bottom quintile of the income distribution from one generation to the next. Hertz (2005) was the first to use transition probabilities to examine black–white differences in intergenerational mobility. Using the Panel Study of Income Dynamics (PSID), Hertz found that blacks were less upwardly mobile and more downwardly mobile over generations than whites. Since then, a few other studies mostly using the PSID have found similar results.[2]

The second set of measures, called directional rank mobility, compares whether the rank of a child in the income distribution is higher or lower than their parents' rank in the previous generation.[3] Both types of measures are able to distinguish upward movements from downward ones and can be measured at different points in the income distribution. The directional rank mobility measure is a useful complement to the transition probability because instead of using an arbitrarily chosen cutoff, it uses a natural yardstick, one's own parents' rank. As I discuss later, the directional rank mobility measures also appear to be very robust to many measurement issues.

A key finding is that in recent decades, blacks have experienced substantially less upward intergenerational mobility and substantially more downward intergenerational mobility than whites. These results are shown to be highly robust to a variety of measurement issues, such as the concept of income used, the age of the sample members, and the length of the time average used. The results are found in two different data sets that cover different birth cohorts and differ in their gender composition. Moreover, these results utilize relatively large samples of black families, so that racial differences can be shown to be statistically significant. An important implication of the results that has not been shown explicitly before is that if these patterns of mobility were to persist into the future, the implications for racial differences in the "steady-state" distribution of income would be alarming. Instead of eventually "regressing to the mean," as some traditional measures of intergenerational mobility (when applied to the whole population) would suggest, these results imply that black Americans would make no further relative progress. Of course, it is a strong hypothetical to assume that current rates of mobility will hold in future generations. Indeed, over the past 150 years, there have been clear periods in which the racial gap in economic status has narrowed and it is certainly possible that black–white gaps could converge.[4]

This study also tries to shed light on which factors are associated with the racial gaps in upward and downward mobility. To be clear, while the analysis is descriptive and not causal, it nonetheless provides some highly suggestive "first-order" clues for the underlying mechanisms leading to black–white differences in intergenerational mobility. It appears that cognitive skills during adolescence, as measured by scores on the Armed Forces Qualification Test (AFQT), are strongly associated with these gaps. For example, conditional on having the median AFQT score, the racial gaps in both upward and downward mobility are relatively small.[5] Consistent with previous studies linking AFQT scores to racial differences in adult outcomes (for example, Neal and Johnson, 1996; Cameron and Heckman, 2001), I do not interpret these scores as measuring innate endowments but rather as reflecting the accumulated differences in family background and other influences that are manifested in test scores.[6] If these results are given a causal interpretation, they suggest that actions that reduce the racial gap in test scores could also reduce the racial gap in intergenerational mobility.[7]

A commonly proposed explanation for racial gaps in achievement has been the relatively high rates of black children growing up with single mothers. I find evidence that for blacks, the lack of two parents in the household throughout childhood does indeed hamper upward mobility. However, patterns in downward mobility are unaffected by family structure for either blacks or whites. Importantly, the negative effects of single motherhood on blacks are only identified in the SIPP, where the entire marital history during the child's life is available. This highlights the importance of access to data on family structure over long periods rather than a single snapshot at one point in time. I also find that black–white gaps in both upward and downward mobility are significantly smaller for those who have completed 16 years of schooling.[8]

In many ways, this work is complementary to the recent study by Chetty et al. (2014) that has deservedly received a great deal of media attention. Chetty et al.

used very large samples of tax records to construct measures of intergenerational mobility at a very detailed level of geography.[9] They then showed how differences in intergenerational mobility across places vary with other aggregate measures, such as the level of segregation or family structure. However, their tax data do not include basic individual characteristics, such as race or education. Therefore, they are unable to show how intergenerational mobility differs by race, which is the first key focus of this article.[10] In addition, they cannot include *individual*-level variables, such as parent education, marital status, wealth, or children's test scores, in order to explain mobility differences, which is the second key focus of this article.

Finally, I should also note that the focus of this article is on *relative* mobility across generations and that the measures are relevant for answering questions concerning the progress of blacks relative to whites. It may also be interesting to consider measures of *absolute* mobility, but that is not the focus of this article.

Measures of mobility

Transition probabilities

The upward transition probability (hereafter UTP) used in this analysis is the probability that the child's income percentile (Y_1) exceeds a given percentile, s, in the child's income distribution by an amount τ, conditional on the parent's income percentile (Y_0) being at or below s in the parent's income distribution:

$$1)\quad UTP_{\tau,s} = \Pr(Y_1 > s + \tau \mid Y_0 \leq s).$$

For example, in a simple case where $\tau = 0$ and $s = 0.2$, the upward transition probability ($UTP_{0,0.2}$) would represent the probability that the child exceeded the bottom quintile in the child's generation, conditional on parent income being in the bottom quintile of the parent generation.[11] The empirical analysis of upward transition probabilities will vary s in increments of 10 percentiles throughout the bottom half of the distribution (that is, 10, 20, … , 50). Therefore, as s increases, each successive sample will add more families to the already existing sample. For example, when $s = 0.1$, only families in the bottom decile of the income distribution will be included. When $s = 0.2$, the sample will now include families in the bottom quintile of the income distribution, so that families in the bottom decile are common to both samples but families who are between the 11th and 20th percentiles are now added. This approach and the use of τ are helpful for making comparisons with the directional mobility estimator that I will introduce shortly. I will also show results that use non-overlapping percentile *intervals* of the parent income distribution (for example, $s \leq$ 10th percentile, 10th percentile $> s \leq$ 20th percentile ... 40th percentile $> s \leq$ 50th percentile). Although in principle the interval-based estimates might be more transparent in pinpointing mobility differences at different points in the distribution, unless one has much larger samples, the results are also much noisier than those from using the cumulative samples.

It is straightforward to see that this estimator can be modified to measure downward transition probabilities by altering the inequality signs:

$$2)\quad DTP_{\tau,s} = \Pr(Y_1 \leq s + \tau \mid Y_0 > s).$$

In this case, I vary s from 50 to 90. I also consider intervals such as the highest decile: 90th percentile $< s \leq$ 100th percentile, next highest decile: 80th percentile $< s \leq$ 90th percentile, … , 50th percentile $< s \leq$ 60th percentile.

Bhattacharya and Mazumder (2011) show how the transition probability can be estimated conditional on continuous explanatory variables using nonparametric regression techniques and demonstrate that bootstrapping is a valid approach for calculating the appropriate standard errors.[12] Using this methodology one can, for example, estimate the difference in transition probabilities between blacks and whites while controlling for the effects of children's test scores and determine whether these differences are statistically significant.

Directional rank mobility

Following Bhattacharya and Mazumder (2011), I use a measure of upward rank mobility (URM) that estimates the likelihood that an individual will surpass their parent's place in the distribution by a given amount, conditional on their parents being at or below a given percentile:

$$3)\quad URM_{\tau,s} = \Pr(Y_1 - Y_0 > \tau \mid Y_0 \leq s).$$

In the simple case where $\tau = 0$, this is simply the probability that the child exceeds the parent's place in the distribution. As with the UTP measure, positive values of τ enable one to measure the *amount* of the gain in percentiles across generations. Results will be shown for a range of values for τ and also as s is progressively increased. Bhattacharya and Mazumder show that the URM measure can also be estimated conditional on continuous covariates using nonparametric regressions.

Similarly, one can construct a measure of downward rank mobility (DRM) using an analogous approach:

4) $$DRM_{\tau,s} = \Pr(Y_0 - Y_1 > \tau \mid Y_0 \geq s).$$

Comparison of transition probabilities and directional rank mobility

One criticism of transition probabilities is that they require using arbitrarily chosen cutoffs such as the 20th percentile. In contrast, the directional rank mobility measures simply compare the child's rank to the parent's rank rather than to an arbitrarily chosen quantile.[13] In other respects, however, neither estimator is perfect, as discussed by Bhattacharya and Mazumder (2011). Therefore, it seems reasonable to consider both measures and to examine a range of estimates. As I show next, the DRM appears to be robust to the differences across data sets.

Data

NLSY79

The first source of data I use is the *National Longitudinal Survey of Youth 1979* cohort (NLSY79), a data set that has several attractive features. Most notably, there is a very large sample of more than 6,000 individuals for whom we know both family income in adolescence (1978–80) and various economic outcomes as adults (1997–2005).

The NLSY79 began with a sample of individuals who were between the ages of 14 and 21 as of January 1, 1979, and who have since been tracked through adulthood. The NLSY79 conducted annual interviews until 1994 and has since shifted to biennial surveys. The analysis is restricted to the sample of youth who were living at home with their parents during the first three years of the survey and for whom family income was directly reported by the parents in any of these years. Respondents also must have stayed in the sample to adulthood and been interviewed in one of the surveys beginning with 1998 and ending in 2006.[14] The final sample includes 3,440 men and 3,250 women.

The measures of mobility utilize data on the family income of the children during the years 1997, 1999, 2001, 2003, and 2005, when sample members were between the ages of 33 and 48. The measures of permanent family income are constructed for each generation by using multiyear averages using *any* available years of data. Years of zero income are included in the averages. Family income is converted into 2004 dollars using the headline Consumer Price Index CPI series.

A nice feature of the NLSY79 is that it also includes a rich set of explanatory variables pertaining to the children. Measures of human capital include completed years of education and scores on the Armed Services Vocational Aptitude Battery test (ASVAB), which was given to all NLSY respondents. I will focus on the composite AFQT score, which is used as a screening device by the military and has been used in many previous economic studies.

SIPP–SSA

The second data source pools the 1984, 1990, 1991, 1992, and 1993 panels of the *Survey of Income and Program Participation* (SIPP) matched to administrative earnings records maintained by the Social Security Administration (SSA).[15] The Census Bureau attempted to collect the social security numbers of all individuals in the surveys and they were subsequently matched to SSA administrative databases of summary earnings records (SER) and detailed earnings records (DER). Davis and Mazumder (2011) show that the match rates are high for most SIPP panels and that selection does not appear to be a serious concern.

The SER data cover annual earnings both from employers and self-employment over the period from 1951 to 2007. In the SER data, the earnings of individuals who are not covered by the social security system will have their earnings recorded as zero. Further, the SER data are censored at the maximum level of earnings subject to the social security tax. While the DER data are not subject to either of these issues, they are only available from 1978. Further, the DER data used in this article only cover labor market earnings reported on W-2 forms and not self-employment earnings. Therefore, I combine information from both the SER and DER by taking the maximum value of earnings from the two sources in order to have earnings data from both labor market earnings and self-employment, and I only use the data beginning in 1978.[16]

In order to maximize the sample size, I use a relatively liberal set of sample selection rules. I start with a sample of white or black males who were living with their parents at the time of the SIPP and who were no older than 25.[17] I also require that the adult earnings of these men are observed when they are at least 21 years old. Sons' earnings are taken over the five years spanning 2003 through 2007, so as to take earnings at as late a stage in the life cycle as possible to minimize life-cycle bias for the younger cohorts. Although years of zero earnings are included in the average, sons must have positive earnings in at least one year to be included. This produces a sample of 16,782 men, who could have been born anytime between 1959 and 1982 and are observed as adults between the ages of 21 and 48.[18]

For children who lived with both their mother and their father, both parents' earnings are combined and averaged over all years between 1978 and 1986 to construct a measure of permanent earnings. For those children who only lived with a single parent at the time of the SIPP, the parent earnings are recorded as the single parent's earnings. To be included in the sample, parents must have had positive earnings in at least one year.

A limitation of the SIPP–SSA data is that there is little information available for the children during their adult years, aside from their administrative earnings records. However, unlike the NLSY, the SIPP–SSA provides a rich set of data on the parents. In this article, I utilize information on the complete marital histories of the parents present at the time of the SIPP.

Comparison of NLSY79 and SIPP–SSA

Table 1 presents summary statistics for each sample. There are a number of potentially important differences between the samples. The NLSY79 sample includes both sons and daughters and uses family income for both generations. Family income is useful as a way of including daughters in the sample and avoiding issues dealing with selective labor force participation. The administrative data in the SIPP–SSA only has earnings and only for the individuals (not the spouse). Since there is no ideal way of dealing with selection of which daughters participate in the labor force, the analysis with the SIPP–SSA only uses sons. The NLSY79 covers individuals born between 1957 and 1964, while the SIPP sample covers those born over a much longer time span, 1959–82.[19] Parent income is measured over just a three-year period (1978 to 1980) in the NLSY79, but over a nine-year period from 1978 to 1986 in the SIPP. All ranks and quantiles used in the NLSY are based on distributions that include individuals who are neither white nor black. The SIPP–SSA data in contrast is restricted to just whites and blacks. Finally, the NLSY will continue to track respondents and collect income data among respondents even if they become incarcerated. The SIPP–SSA data, in contrast, is confined to the civilian noninstitutionalized population.

Unconditional estimates of intergenerational mobility

Upward transition probabilities (UTP)

The race-specific estimates of upward transition probabilities from both data sets are plotted in figure 1.[20] The x-axis varies the sample used based on the percentile range of family income in the parent generation, while the y-axis shows the transition probability that income of children from these families surpassed this range. The blue lines show the estimates for whites, while the red lines show the estimates for blacks. The green lines plot the difference in the probabilities by race, along with standard error bands.[21] The solid lines show the estimates using the NLSY and the dashed lines use the SIPP–SSA sample.

I begin by discussing the results using the NLSY. Among white men and women (solid blue line) whose parents' income was at or below the 10th percentile, 84 percent exceed the 10th percentile as adults. As we move to the right and gradually increase the percentile range of family income, the upward transition probabilities fall. For example, among whites starting below the 40th percentile in the parent generation only 54 percent exceed the 40th percentiles as adults. In all cases, the comparable UTP estimates are much lower among blacks (solid red line). For example, among blacks starting in the bottom decile, only 65 percent exceed the bottom decile as adults, a 19 percentage point difference compared with whites. The black–white gap in the probability of rising out of the bottom quintile (solid green line) is even higher at 27 percent.

The SIPP–SSA sample consists only of sons, includes only blacks and whites, includes many more recent cohorts, and uses administrative earnings data rather than family income. Despite these different concepts and measures, the UTP estimates are very similar to those using the NLSY. The general pattern of large and statistically significant differences in point estimates is also evident in the SIPP–SSA data. The fact that the key findings are so similar across the data sets is advantageous, since each data set has its own exclusive set of explanatory variables.

Downward transition probabilities (DTP)

Figure 2 (p. 8) plots an analogous set of downward transition probabilities. Using either data set, I find that blacks are more downwardly mobile. This is most evident when the sample includes a broad range of the upper income distribution in the parent generation. For example, about 60 percent of blacks whose parents were in the top half of the income distribution fall below the 50th percentile in the subsequent generation. The analogous figure for whites is 36 percent.[22]

Upward rank mobility (URM)

Figure 3 (p. 9) plots the estimates of upward rank mobility based on equation 3. As might be expected, the rates of upward mobility using the URM are somewhat higher than for the UTP. For example, using the NLSY I find that 75 percent of blacks whose parents were below the 20th percentile surpass their parents' percentile in the family income distribution. In contrast, only 48 percent of this same subsample exceed the 20th percentile, implying that although about 37 percent of blacks starting in the bottom quintile exceed their parents' percentile but do not transition out of the bottom quintile. For whites, the difference in upward mobility

TABLE 1

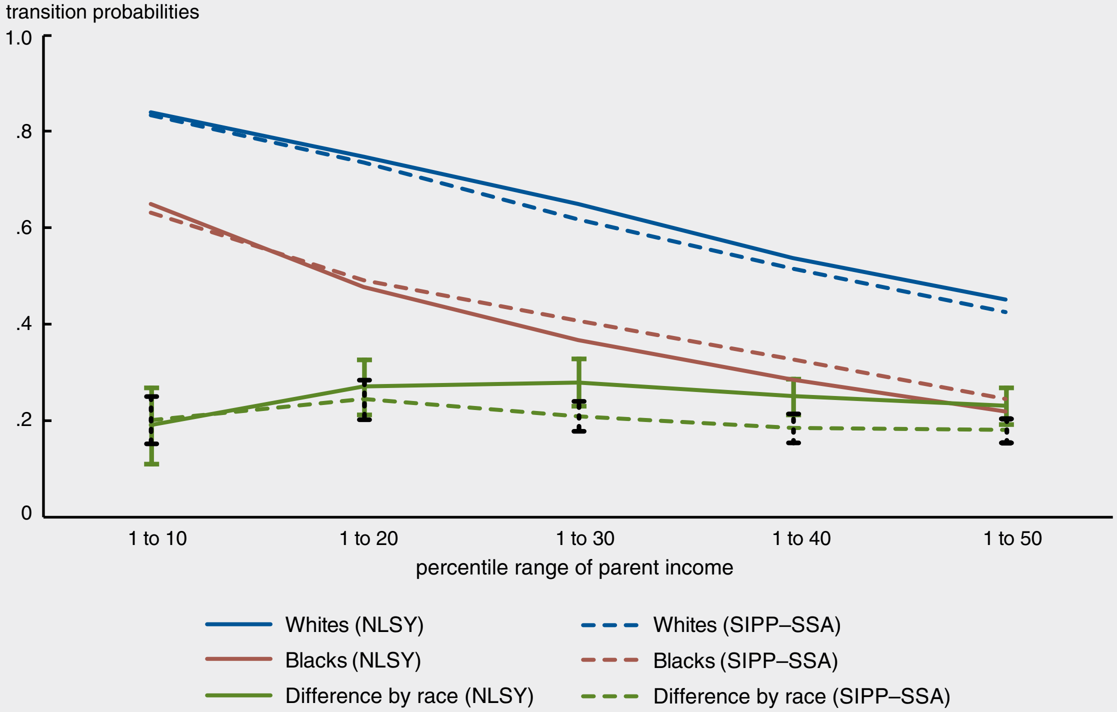
Summary statistics

	All			Whites			Blacks		
	Number of observations	Mean	Standard deviation	Number of observations	Mean	Standard deviation	Number of observations	Mean	Standard deviation
Panel A: NLSY79									
Family income (1997–2005)	6,690	69,395	58,953	3,205	76,284	61,316	2,143	42,289	39,070
Child's age in 2001	6,690	39.1	2.2	3,205	39.1	2.2	2,143	39.2	2.2
Education	6,673	13.0	2.3	3,199	13.2	2.3	2,136	12.4	2.1
AFQT	6,432	46.0	28.5	3,080	52.5	27.0	2,082	21.6	19.8
Parent income (1978–80)	6,690	57,760	36,299	3,205	64,354	35,965	2,143	33,743	26,725
Single mother at age 14	6,690	0.13	0.33	3,205	0.08	0.27	2,143	0.3	0.5
Panel B: SIPP–SSA									
Son's log earnings	16,782	10.14	1.07	14,757	10.23	1.00	2,025	9.48	1.32
Son's age in 2005	16,782	30.93	5.69	14,757	30.97	5.71	2,025	30.66	5.55
Single parent	16,782	0.21	0.41	14,757	0.17	0.38	2,025	0.53	0.50
1984 panel	16,782	0.26	0.44	14,757	0.25	0.43	2,025	0.27	0.45
1990 panel	16,782	0.23	0.42	14,757	0.23	0.42	2,025	0.23	0.42
1991 panel	16,782	0.14	0.35	14,757	0.15	0.35	2,025	0.12	0.33
1992 panel	16,782	0.20	0.40	14,757	0.20	0.40	2,025	0.19	0.40
1993 panel	16,782	0.18	0.38	14,757	0.18	0.38	2,025	0.18	0.38
Parent log earnings	16,782	10.32	1.06	14,757	10.42	1.00	2,025	9.63	1.26
Father's age in 1982	15,354	35.72	8.85	13,467	36.09	8.75	1,887	33.11	9.14

Notes: NLSY79 is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth 1979*. AFQT is the Armed Forces Qualification Test. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

FIGURE 1

Upward transition probabilities by race using cumulative samples ($\tau = 0$)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

between the two measures is much smaller. Therefore, the URM estimator (for $\tau = 0$) shows a much smaller black–white gap that fluctuates around 0.1 across the samples in the bottom half of the income distribution of parents. Interestingly, figure 3 shows that the estimates based on the URM are nearly identical across the two data sets, which suggests that it is an especially robust measure.

The finding of a smaller black–white gap using the URM rather than the UTP measure is sensitive to the chosen value of τ. For example, if τ is set to 0.2, then the black–white differences in upward rank mobility rise considerably. For example, among men and women in the NLSY whose parents' family income placed them in the bottom quintile, blacks are nearly 25 percent less likely to surpass their parents' rank by 20 percentiles or more. Using the SIPP–SSA data, the analogous black–white difference for men is 21 percent. Figure 4 (p. 10) plots the full set of estimates for the case where τ equals 0.2.

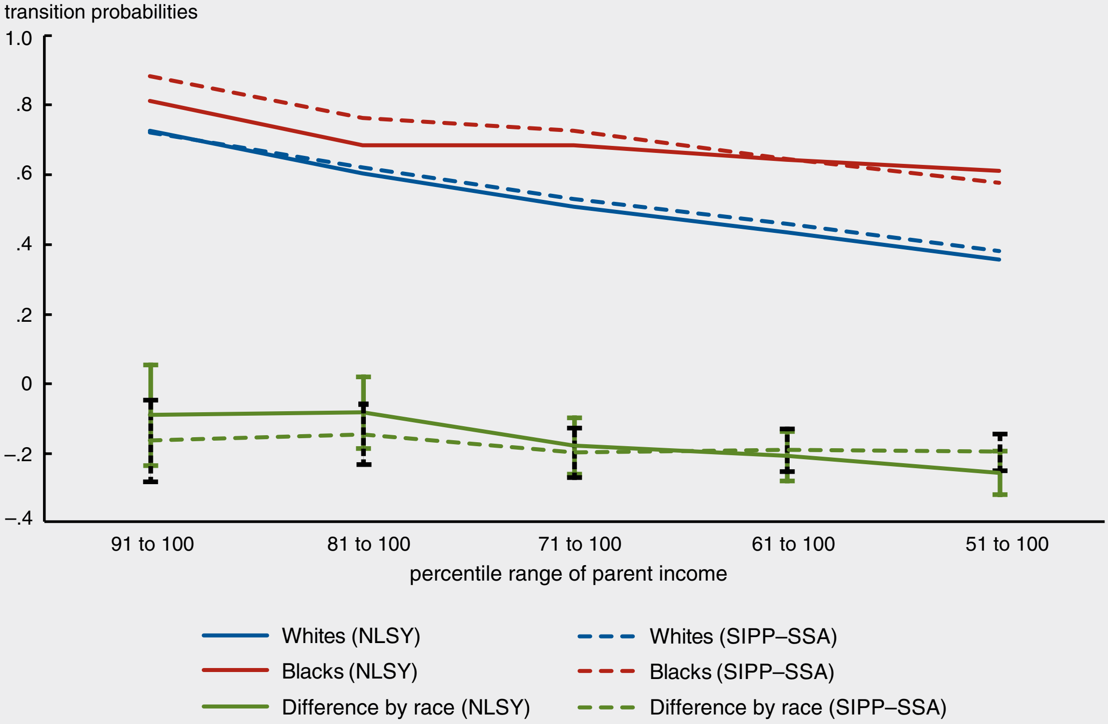
Downward rank mobility (DRM)

Estimates of downward rank mobility are shown in figure 5 (p. 11). Using the simple measure ($\tau = 0$), I again observe higher rates of downward mobility among blacks than whites that is less pronounced in the top two deciles. The estimates of DRM are higher, however, than those of DTP. For example, among whites in the NLSY sample whose parents' income was in the top half of the income distribution, 69 percent were in a lower rank in the distribution than their parents, even though only 36 percent fell below the median. For blacks starting in the top half of the income distribution, 79 percent fell below their parents and 61 percent also dropped below the median. Therefore, the estimates of the black–white gap in downward mobility using the baseline DRM measure are considerably smaller in absolute value than the analogous estimates using DTP.

The comparison of the two downward mobility measures is also sensitive to the choice of τ. For example, if we consider the probability of those in the top half of the distribution falling 20 percentiles or more, the black–white gap is 18 percent in the NLSY

FIGURE 2

Downward transition probabilities by race using cumulative samples (τ = 0)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

and 14 percent in the SIPP–SSA. The racial differences in DRM when τ = 0.2 show somewhat different patterns across the income distribution depending on the data set used, as shown in figure 6 (p. 12). For example, the black–white difference in the probability of falling 20 percentiles below one's parents among those who start in the top decile is only 7 percent in the NLSY, but it is 23 percent in the SIPP–SSA. This likely reflects differences that are due to the relevant concept of income. Compared with whites, blacks starting in the top decile are more likely to suffer larger drops in their earnings rank than in their family income rank.

Upward mobility using interval-based samples

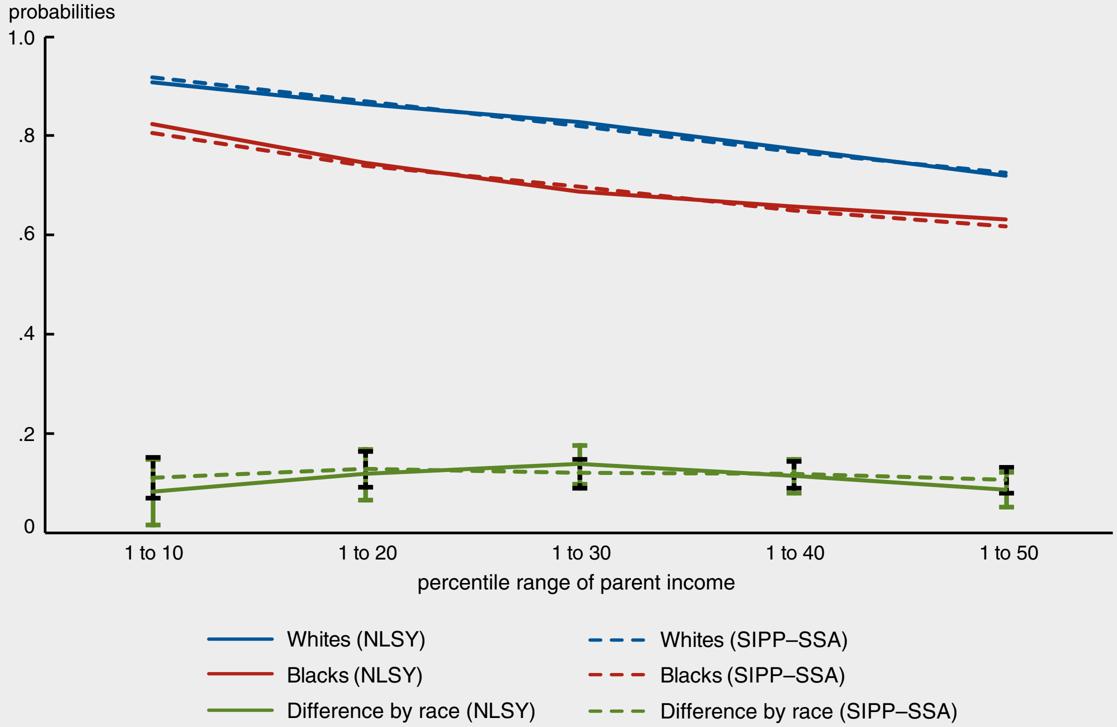
The results presented so far have used samples that have progressively expanded the range of families, starting from either the bottom or the top of the parent income distribution. It might also be interesting to see the results within much narrower percentile ranges to see how mobility changes across the distribution. Figure 7 (p. 13) shows estimates of UTP and URM using interval-based samples from deciles in the bottom half of the income distribution.[23] For most of the bottom half of the income distribution, the racial differences in upward mobility are consistently between 20 and 30 percent. The greater similarity between the UTP and URM estimates is not surprising since, as the interval range becomes smaller, the two estimates will converge.[24]

Implications of transition probabilities on the steady-state distributions by race

The transition matrix of movements across quintiles of the income distribution over generations for blacks and whites based on the SIPP–SSA are shown in table 2 (p. 14). The general patterns concerning racial differences in upward and downward mobility are again evident. For example, more than 50 percent of blacks who start in the bottom quintile in the parent generation remain there in the child generation, but only 26 percent of whites remain in the bottom quintile in both generations. Whites are less likely to transition out of the top quintile compared with blacks, suggesting a distribution that may not be exhibiting racial convergence. Assuming that these specific probabilities are a

FIGURE 3

Upward rank mobility by race using cumulative samples ($\tau = 0$)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

permanent feature of the U.S. economy, they can be used to calculate an implied steady-state distribution using standard matrix algebra methods for solving Markov chains. The results show, for example, that in the steady state, 39 percent of blacks would occupy the bottom quintile of the income distribution and only 8 percent would be in the top quintile.[25] This finding suggests that rather than convergence, blacks will remain perpetually disadvantaged in American society if mobility patterns continue to evolve as they have for the cohorts studied in this article. As I discuss next, however, there are potential levers through which policy could address this problem.

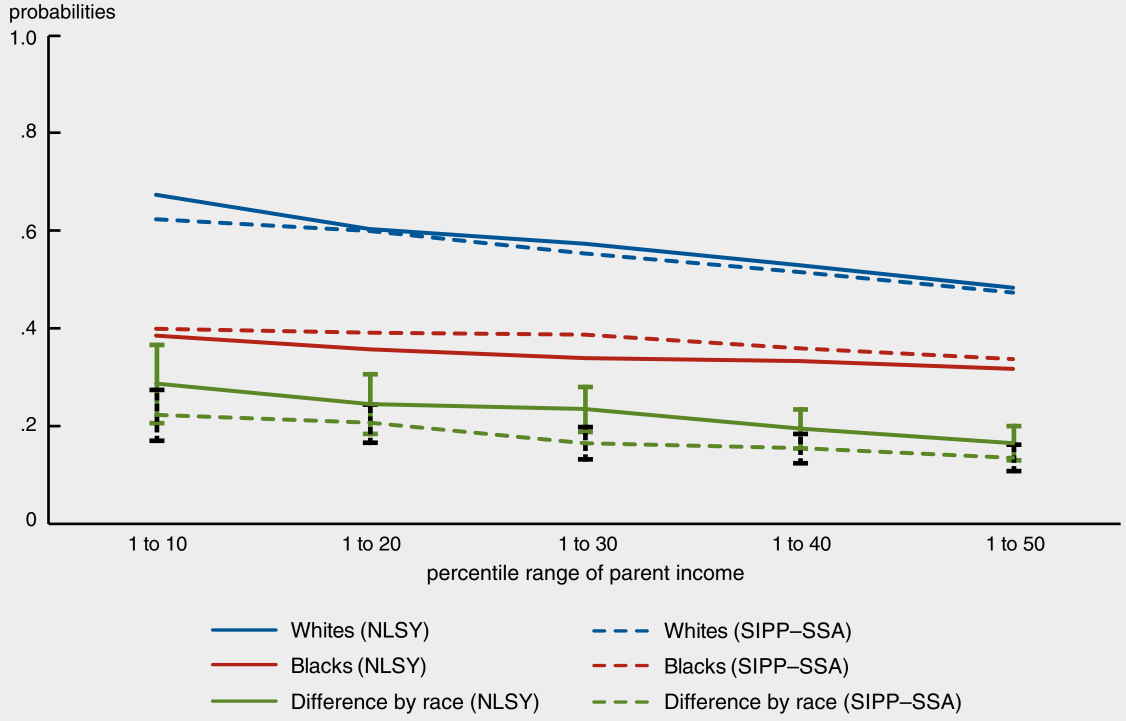
Estimates of intergenerational mobility controlling for explanatory variables

Ideally, we would like to understand the causal factors that explain the observed patterns of intergenerational mobility and the possible implications for policies designed to address racial differences in mobility. For example, we might like to know whether a particular schooling intervention such as smaller classes might improve students' prospects for upward mobility and whether this could reduce the racial gap in upward mobility. Such a study would not only require a convincing research design to address standard concerns about endogeneity bias—for example, to ensure that the intervention was not directed at individuals who would have succeeded even in the absence of the intervention—but would also likely require high-quality income data spanning multiple years of adulthood for *two generations of the same set of families*. Instead, like the recent work by Chetty et al. (2014), which also does not attempt to estimate causal effects, I opt for a more modest goal and conduct a descriptive analysis to explore how the inclusion of other available explanatory variables affects the racial differences in upward and downward intergenerational mobility. Such a "first pass" analysis may yield useful clues about which factors are potentially important.

To simplify the analysis, I focus only on transition probabilities.[26] For a representative measure of upward mobility, I use the transition probability of moving out of the bottom quintile. For downward mobility,

FIGURE 4

Upward rank mobility by race using cumulative samples ($\tau = 0.2$)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

I focus only on the probability of moving out of the top half of the income distribution over the course of a generation. I first consider the effects of two explanatory variables from the NLSY, the child's education level and the child's test score. I then turn to a measure of family structure from the SIPP–SSA data, which compares children who have ever lived with just one parent with children who have always lived with two parents.

For the first two measures, I use a statistical technique that shows how each explanatory variable affects whether a child exceeds the bottom quintile as an adult and how this association changes at different values of the explanatory variable.[27] I produce a series of plots of the upward transition probability at each value of the explanatory variable for each racial group. In addition, I plot the black–white difference, along with 95 percent confidence bands. Finally, as a point of reference, I include the baseline transition probabilities that do not account for the explanatory variables in lightly shaded horizontal lines. An explanatory variable with a positive association with upward mobility will produce an upward sloped line and may reduce the black–white gap in upward mobility.

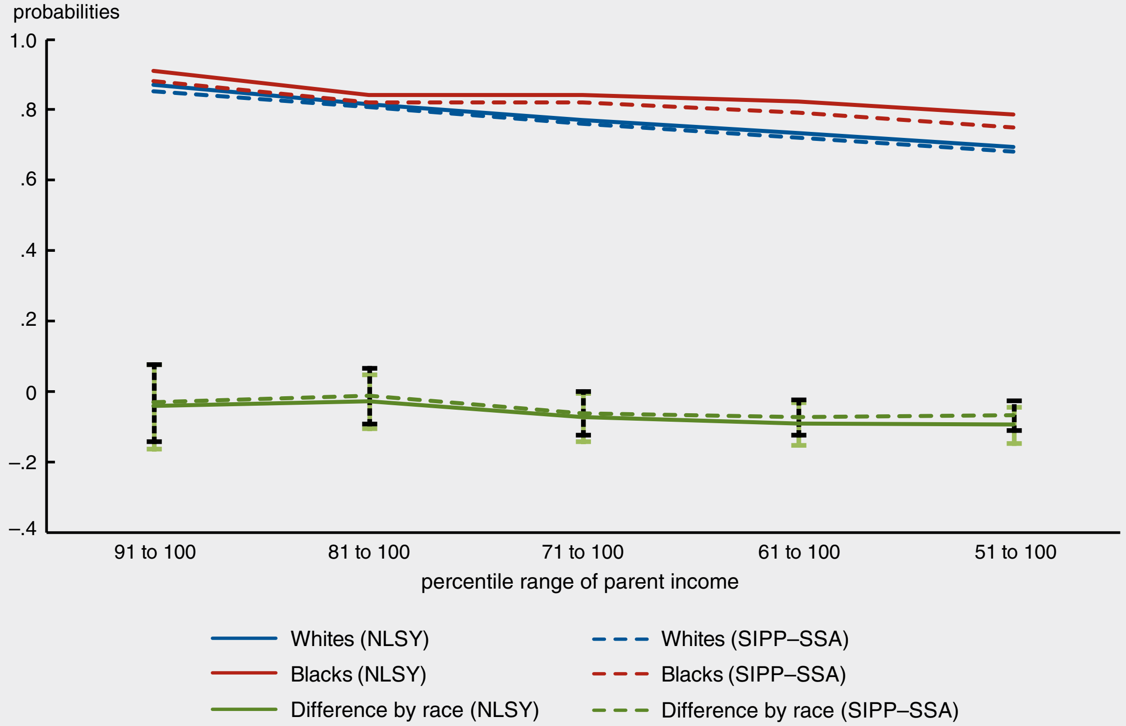
Effects of education

The left-hand-side panels in figure 8 (p. 15) show the results for upward mobility and the right-hand-side panels show the plots for downward mobility. Panel A shows that, as would be expected, more years of completed schooling are associated with a greater likelihood of rising out of the bottom quintile. For example, 89 percent of whites with exactly 16 years of schooling will escape the bottom quintile, compared with 75 percent of whites with exactly 12 years of schooling. For blacks, rates of upward mobility are extremely low for those with less than a high school education but begin to rise sharply for those who attain more than a high school education. For example, for blacks with exactly ten years of schooling, only 28 percent will transition out of the bottom quintile, compared with 69 percent of blacks with exactly 14 years of schooling.

With respect to the *racial gap* in upward mobility, controlling for education provides something of a mixed

FIGURE 5

Downward rank mobility by race using cumulative samples (τ = 0)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

picture. On one hand, the point estimate for the racial gap in upward mobility among those with less than a high school education is actually higher than the estimate when education is not controlled for, although this difference is not statistically significant. On the other hand, the racial gap narrows sharply with additional years of post-secondary education. Indeed, among those with 16 years of schooling, the racial gap in upward mobility gap is essentially closed. Nevertheless, the racial gap is still quite large among those with some post-secondary education who have not completed college. For example, the black–white gap among those with 14 years of schooling is still sizable at 16 percent. Given that only 17 percent of blacks in the NLSY attained more than 14 years of schooling, this suggests that *marginal* improvements in educational attainment may not do a great deal to improve the overall upward mobility prospects of blacks.

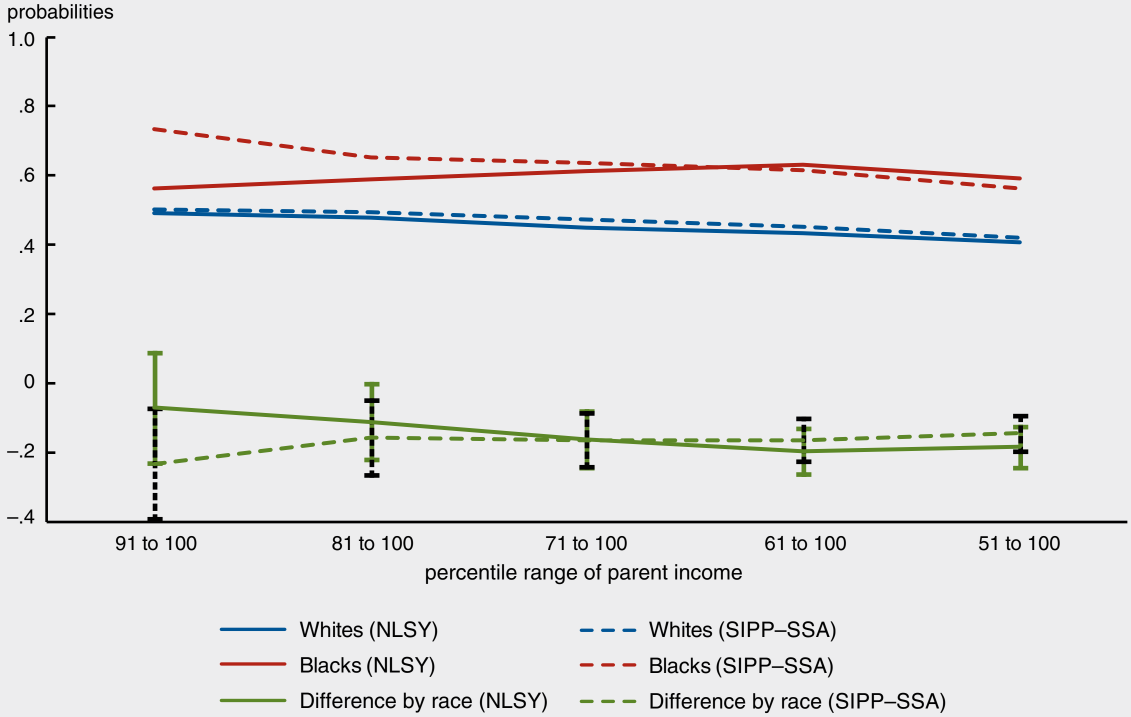
The effects of education on downward mobility are shown in panel B of figure 8. As expected, the lines slope downward. Since I am sampling only families with parents in the top half of the income distribution, the samples of individuals with less than a high school education are relatively small, so the estimates for these values are especially noisy. As was the case with upward mobility, additional years of post-secondary schooling are associated with a reduction in the racial gap in downward mobility. Among those with 16 years of schooling, the black–white gap is reduced to just 14 percentage points, and it disappears entirely among those with 17 years or more of schooling.

Effects of test scores

The effects of including one's AFQT score on rates of upward mobility are shown in panel C of figure 8. Here, the results provide a relatively clean and compelling story. For both blacks and whites, upward mobility rises with AFQT scores in a fairly similar fashion. There are especially sharp gains in upward mobility associated with increases in test scores at the low end of the AFQT distribution. Upward mobility continues to rise at a somewhat slower but still strong rate in the middle and in the upper half of the AFQT distribution. Remarkably, the lines for blacks

FIGURE 6

Downward rank mobility by race using cumulative samples (τ = 0.2)



Notes: NLSY is the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth*. SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

and whites are relatively close throughout the AFQT distribution. For example, the black–white gap in moving out of the bottom quintile is only 5.2 percentage points for those with median AFQT scores, compared with the unconditional gap of 27 percentage points. This suggests that cognitive skills measured at adolescence can "account" for much of the black–white difference in upward mobility. This result echoes previous findings by Neal and Johnson (1996) and Cameron and Heckman (2001), who have also found that AFQT scores can account for much of the racial gap in adult earnings and college enrollment rates. As with these aforementioned studies, I interpret this finding as reflecting the cumulative effect of a broad range of family background influences rather than reflecting only innate differences.[28]

The effects of AFQT scores on downward mobility (panel D of figure 8) are also quite striking. The lines for whites and blacks converge quite a bit and for a broad swath of the AFQT distribution, the racial gap is below 10 percentage points and is not statistically different from 0. Therefore, as was the case with upward mobility, test scores during adolescence are strongly associated with rates of downward mobility.

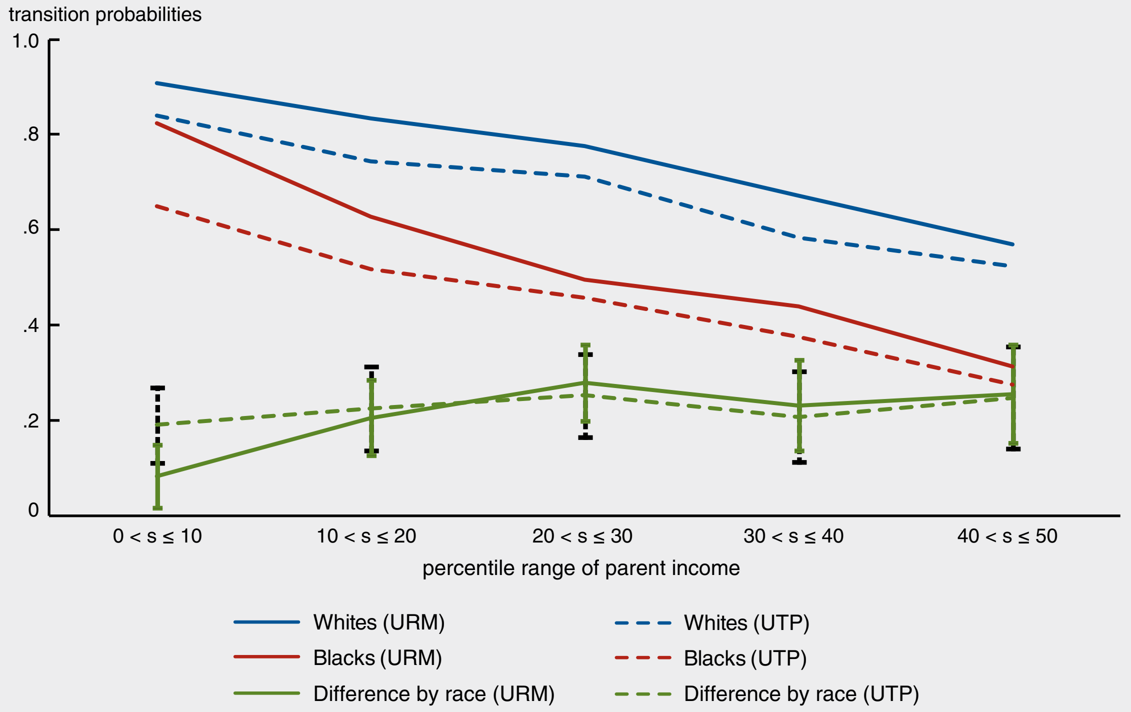
Effects of family structure

To understand the role of family structure, I use data from the SIPP–SSA where I have data on family structure throughout the child's life.[29] For whites, upward mobility out of the bottom quintile actually declines slightly from 0.75 (0.02) for those who ever lived with just one parent to 0.71 (0.02) for those who always lived with two parents. For blacks, however, we see an increase in the transition probability from 0.47 (0.02) to 0.58 (0.02). The black–white gap declines from 0.28 (0.02) to 0.13 (0.06). This 15 percentage point improvement in upward mobility for blacks relative to whites is statistically significant at the 5 percent level. On the other hand, I find virtually no difference in downward mobility by whether sons always lived with two parents or not, for either blacks or whites.

It is worth noting that Chetty et al. (2014) show that the strongest predictor of intergenerational mobility differences between commuting zones is the fraction

FIGURE 7

Upward mobility estimates by race using intervalled samples ($\tau = 0$)



Note: URM indicates upward rank mobility. UTP indicates upward transition probabilities.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

of families in the commuting zone headed by single mothers. However, Chetty et al. do not show the direct effect of having a single parent on intergenerational mobility differences at the individual level.

Conclusion

One can potentially gain insight into the dynamics of the racial gap in economic status in the United States and better understand how long it will take before there is complete convergence by examining rates of intergenerational income mobility. Using measures that are suited to describing racial differences in intergenerational mobility with respect to a common distribution, I find dramatically lower rates of upward mobility from the bottom of the income distribution and dramatically higher rates of downward mobility from the top of the distribution among blacks born between the late 1950s and early 1980s.

In combination the estimates imply a steady-state income distribution that shows no racial convergence. In other words, if future generations of white and black Americans experience the same rates of intergenerational mobility as these cohorts, we should expect to see that blacks on average would not make any relative progress. While these results are provocative, they stand in contrast to other epochs in which blacks have made steady progress in reducing racial differentials. These findings, therefore, should not be taken to imply that racial progress is infeasible but rather to highlight what current trends, if they were to continue, would suggest about the future.

These results also underscore the importance of understanding what kinds of policies can potentially foster greater upward mobility and reduce downward mobility for blacks. While this article does not seek to identify definitive causal channels, the use of statistical models that include explanatory variables suggests a few potential areas for policymakers to consider. Similar to previous studies that have looked at static gaps in black–white earnings and college-going rates using NLSY data (for example, Neal and Johnson, 1996; Cameron and Heckman, 2001), it is apparent that the cumulative effects of a variety of influences that affect cognitive ability by adolescence play a critical role in

TABLE 2

Transition matrices by race using SIPP–SSA sample

	Child's income quintile				
Parents' income quintile	1	2	3	4	5
Panel A: Whites					
1	**0.263**	**0.267**	**0.208**	**0.159**	**0.103**
	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)
	2,510	2,510	2,510	2,510	2,510
2	**0.205**	**0.239**	**0.219**	**0.204**	**0.133**
	(0.009)	(0.008)	(0.007)	(0.006)	(0.005)
	2,815	2,815	2,815	2,815	2,815
3	**0.156**	**0.203**	**0.236**	**0.223**	**0.182**
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
	2,999	2,999	2,999	2,999	2,999
4	**0.147**	**0.162**	**0.206**	**0.234**	**0.250**
	(0.007)	(0.006)	(0.007)	(0.006)	(0.006)
	3,165	3,165	3,165	3,165	3,165
5	**0.113**	**0.136**	**0.155**	**0.217**	**0.380**
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
	3,268	3,268	3,268	3,268	3,268
Panel B: Blacks					
1	**0.508**	**0.207**	**0.155**	**0.092**	**0.038**
	(0.017)	(0.021)	(0.025)	(0.034)	(0.048)
	846	846	846	846	846
2	**0.357**	**0.246**	**0.203**	**0.129**	**0.065**
	(0.013)	(0.020)	(0.023)	(0.028)	(0.041)
	541	541	541	541	541
3	**0.341**	**0.212**	**0.176**	**0.190**	**0.081**
	(0.012)	(0.018)	(0.021)	(0.030)	(0.045)
	358	358	358	358	358
4	**0.272**	**0.236**	**0.173**	**0.178**	**0.141**
	(0.010)	(0.012)	(0.019)	(0.028)	(0.042)
	191	191	191	191	191
5	**0.213**	**0.180**	**0.180**	**0.191**	**0.236**
	(0.007)	(0.011)	(0.015)	(0.026)	(0.048)
	89	89	89	89	89

Notes: SIPP–SSA matches data from the U.S. Census Bureau's *Survey of Income and Program Participation* to administrative earnings records from the Social Security Administration. Both panels use subsamples drawn from a sample of 16,782 men from the SIPP–SSA data and a multiyear average of sons' earnings over 2003–07 and parents' earnings over 1978–86. Bootstrapped standard errors are in parentheses. Sample sizes are shown below the standard errors.

Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

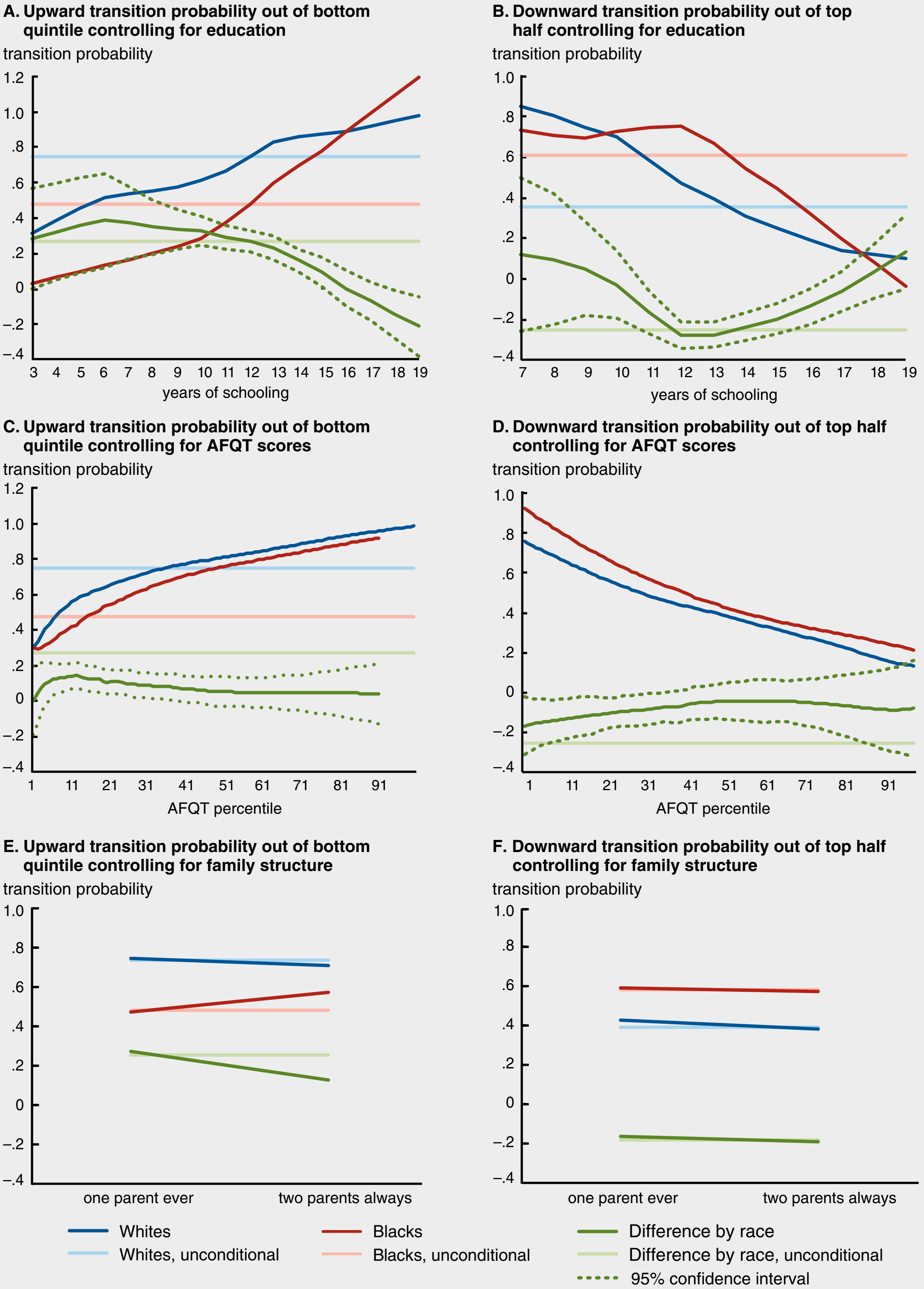
accounting for racial differences in upward and downward mobility. A growing literature has shown that black–white differences in test scores, and military test scores in particular, have been narrowed through large-scale policy interventions throughout American history (for example, Chay, Guryan, and Mazumder, 2009; Aaronson and Mazumder, 2011). Other studies (for example, Dobbie and Fryer, 2011) have also shown the potential for modern educational interventions to improve the black–white gap in educational achievement.

Educational attainment also appears to matter for both upward and downward mobility, but the effects of education on reducing racial mobility differentials occur primarily at the margin of acquiring higher education. If racial gaps in college attainment are primarily due to skill differences determined in adolescence (Cameron and Heckman, 2001), then this also points to the importance of interventions earlier in life. Still, there may be some scope for higher education policies that ease credit constraints for families for whom those constraints are binding.

Many commentators have pointed to the prevalence of black children raised by single mothers as a source of racial gaps in economic success. I find supportive evidence that blacks raised in two-parent families throughout childhood experience significantly greater upward mobility. However, family structure appears not to matter for whites or for rates of downward mobility for either blacks or whites. Future research may provide greater insight into these patterns of results.

FIGURE 8

Transition probability estimates controlling for explanatory variables



Note: AFQT is the Armed Forces Qualification Test.
Sources: Author's calculations based on data from U.S. Bureau of Labor Statistics, U.S. Census Bureau, and U.S. Social Security Administration.

NOTES

[1]President Obama highlighted equality of opportunity in an address in 2011. See www.whitehouse.gov/the-press-office/2011/12/06/remarks-president-economy-osawatomie-kansas. Republican leaders have also raised concerns about low economic mobility. See http://business.time.com/2012/01/05/the-loss-of-upward-mobility-in-the-u-s/.

[2]These include Isaacs, Sawhill, and Haskins (2008), Mazumder (2008), and Acs (2011), which were all produced as part of the Pew Economic Mobility Project. Mazumder (2008) uses the NLSY, while the other studies use the PSID.

[3]In a methodological paper, Bhattacharya and Mazumder (2011) introduced these measures and demonstrated their properties. This article builds upon Bhattacharya and Mazumder in several ways, including adding a second data source that utilizes administrative records; adding daughters' outcomes; considering racial differences in downward mobility; and adding many more explanatory variables to the analysis.

[4]Smith and Welch (1989) show that there was significant convergence in the black–white wage gap from 1940 to 1980 that was due to improvements in black educational attainment and school quality and migration patterns.

[5]This was first found by Bhattacharya and Mazumder (2011). The residual racial gap in upward mobility conditional on AFQT, however, is higher in this study and may be due to the fact that this paper uses children's family income rather than earnings, includes women, and uses income measured at later ages.

[6]A growing literature suggests that black–white differences in test scores can be strongly affected by environmental influences. For example, see Chay, Guryan, and Mazumder (2009) and Aaronson and Mazumder (2011).

[7]There may be other race-specific behavioral differences that can affect the interpretation of economic gaps that adjust only for AFQT scores. Lang and Manove (2011) argue that for signaling reasons, blacks obtain more education than whites conditional on AFQT scores.

[8]In the working paper version of this article (Mazumder, 2011), I show that low levels of parental wealth among blacks also inhibit the prospects for upward mobility.

[9]Chetty et al. (2014) measure intergenerational mobility at the level of "commuting zones," which are an aggregation of counties that include rural areas.

[10]They conduct one exercise using samples that vary in the share of the population that is white (inferred based on geographic residence) in order to show that segregation is associated with reduced upward mobility of whites as well as blacks.

[11]Many previous papers on intergenerational mobility have used a quintile transition matrix to characterize mobility by showing what proportion of those who start in each quintile end up in each quintile in the subsequent generation. This example would measure 1 minus the probability of remaining in the bottom quintile. A quintile transition matrix is shown in table 2.

[12]Formby, Smith, and Zheng (2004) develop a distribution theory for marginal transition probabilities that can be easily extended to the case of discrete covariates. Unfortunately, for many covariates of interest that are commonly treated as continuous, such as years of schooling or test scores, this is not of much practical value. In order to implement the TP estimator, one must first estimate quantiles of the income distribution. Since the TP estimates conditional on continuous covariates will involve non-smooth functions of these initially estimated functions, it is technically challenging to show that one can bootstrap the standard errors.

[13]When making comparisons between population subgroups, there is an unambiguous advantage to using the URM. However, Bhattacharya and Mazumder (2011) show that when using the full sample (that is, pooling all subgroups), the URM measure is only meaningful if there is some cutoff, s, used to condition the sample. The choice of s, of course, is likely to be arbitrary. Even in this case, however, children's ranks are still directly compared to their parents' rank as opposed to an arbitrary quantile.

[14]The analysis includes individuals from both the cross-sectional representative samples and the supplemental samples (for example, blacks and Hispanics). Following Neal and Johnson (1996) and Cameron and Heckman (2001), I combine the cross-sectional and supplemental samples of blacks and utilize the 1979 sampling weights.

[15]This data source is not publicly available. Researchers must apply to obtain the data through the Center for Economic Studies at the U.S. Census Bureau (www.census.gov/ces).

[16]For a small set of self-employed individuals whose earnings were above the taxable maximum, this approach understates their true earnings. To address this, I obtained the full DER data (including the non-top-coded self-employed earnings) and redid all of the analysis. I found that using the full DER data has an imperceptible effect on the results (typically only changing estimates at the third decimal place). Since there are procedural difficulties in releasing a second set of statistical results through the Census Bureau disclosure avoidance review process in cases where revised estimates lead the sample size to change by just one or two individuals, and since the current results are virtually identical to the corrected ones, I have opted to show the current results that combine both the SER and DER data.

[17]Restricting the sample to whites and blacks avoids implicitly disclosing any information concerning men who are neither white nor black, thereby making it easier to pass Census Bureau disclosure avoidance review. Similarly, the upper age restriction of 25 is not ideal—one might prefer a younger age cutoff, such as 17 or 18, to avoid including men who lived at home into adulthood—but I chose it to make it easier to maximize sample sizes for the purposes of Census Bureau disclosure avoidance review. It is worth noting that in the SIPP individuals residing at college are included in the household. To some degree, this mitigates the concern of having an older age cutoff. In any case, the results are not sensitive to restricting the age cutoff to 18. There is no lower bound on the age when living at home.

[18]As I discuss later, the results are not sensitive to requiring sons to be at least 28 years old.

[19]Haider and Solon (2006) demonstrate that estimates of intergenerational elasticity can be biased depending on the ages at which the incomes of children and parents are measured. They find that such bias is minimized when income is measured around the age of 40. It is not clear whether a similar bias would arise with respect to the statistical measures utilized here, since they are very different from the regression coefficients analyzed by Haider and Solon. Indeed, the directional rank mobility measures, in particular, seem to be robust to the various measurement differences between the samples. In any case, life-cycle bias is likely to be minimal in the NLSY79 sample, since the mean age of the children in 2001 (the middle year of the sample) is 39, which is close to ideal according to

Haider and Solon. In the SIPP–SSA sample, the mean age of the sons in 2005 (the middle year of the sample) is 31. However, I found very similar results using an older SIPP–SSA sample.

[20]The underlying estimates and the standard errors are available in tabular form in Mazumder (2011), which is a working paper version of this article.

[21]Owing to the large samples, all of the estimated gaps are highly statistically significant.

[22]There is a somewhat notable difference between the two data sets in the degree of downward mobility out of the top decile for blacks. In the NLSY, which uses family income in both generations, 81 percent of black children whose parents were in the top decile fall below the top decile as adults. The comparable figure is 88 percent in the SIPP–SER data, where the income concept is earnings.

[23]The UTP estimates are drawn from the NLSY sample, while the URM estimates are drawn from the SIPP–SSA sample.

[24]This is obvious at the limit, since the probability of exceeding the percentile (URM) of one's parent(s) and the probability of exceeding any given percentile threshold (UTP) will be identical if the sample is conditioned on the same percentile in each case.

[25]Further, 22 percent of blacks would be in the second quintile, 17 percent in the third quintile, and 14 percent in the fourth quintile. The shares of whites across the distribution from the bottom to top quintiles is as follows: 17 percent, 20 percent, 20 percent, 21 percent, and 22 percent.

[26]I found the same general pattern of results using the DRM measures.

[27]Following Bhattacharya and Mazumder (2011), I use locally weighted regressions and calculate standard errors using the bootstrap method.

[28]A growing number of studies (Neal and Johnson, 1996; Hansen, Heckman, and Mullen, 2004; Cascio and Lewis, 2006; Chay, Guryan, and Mazumder, 2009; and Aaronson and Mazumder, 2011) have shown that environmental influences can have large effects on military test scores and narrow racial differences.

[29]The latter category includes those whose parents were ever separated, divorced, or widowed. The sample includes those who were given the marital history topical module in the SIPP and who had non-missing data. There are no significant differences between this subsample and the full SIPP analysis sample.

REFERENCES

Aaronson, Daniel, and Bhashkar Mazumder, 2011, "The impact of Rosenwald schools on black achievement," *Journal of Political Economy*, Vol. 119, No. 5, October, pp. 821–888.

Acs, Gregory, 2011, "Downward mobility from the middle class: Waking up from the American Dream," Pew Charitable Trusts, Economic Mobility Project, report, September.

Bhattacharya, Debopam, and Bhashkar Mazumder, 2011, "A nonparametric analysis of black–white differences in intergenerational income mobility in the United States," *Quantitative Economics*, Vol. 2, No. 3, November, pp. 335–379.

Cameron, Stephen V., and James J. Heckman, 2001, "The dynamics of educational attainment for black, Hispanic, and white males," *Journal of Political Economy*, Vol. 109, No. 3, June, pp. 455–499.

Cascio, Elizabeth U., and Ethan G. Lewis, 2006, "Schooling and the Armed Forces Qualifying Test: Evidence from school-entry laws," *Journal of Human Resources*, Vol. 41, No. 2, Spring, pp. 294–318.

Chay, Kenneth Y., Jonathan Guryan, and Bhashkar Mazumder, 2009, "Birth cohort and the black-white achievement gap: The roles of access and health soon after birth," National Bureau of Economic Research, working paper, No. 15078, June.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, 2014, "Is the United States still a land of opportunity? Recent trends in intergenerational mobility," National Bureau of Economic Research, working paper, No. 19844, January.

Davis, Jonathan M. V., and Bhashkar Mazumder, 2011, "An analysis of sample selection and the reliability of using short-term earnings averages in SIPP-SSA matched data," U.S. Census Bureau, Center for Economic Studies, working paper, No. CES-WP-11-39, December.

Dobbie, Will, and Roland G. Fryer, 2011, "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, Vol. 3, No. 3, July, pp. 158–187.

Formby, John P., W. James Smith, and Buhong Zheng, 2004, "Mobility measurement, transition matrices and statistical inference," *Journal of Econometrics*, Vol. 120, No. 1, May, pp. 181–205.

Haider, Steven, and Gary Solon, 2006, "Life-cycle variation in the association between current and lifetime earnings," *American Economic Review*, Vol. 96, No. 4, September, pp. 1308–1320.

Hansen, Karsten T., James J. Heckman, and Kathleen J. Mullen, 2004, "The effect of schooling and ability on achievement test scores," *Journal of Econometrics*, Vol. 121, Nos. 1–2, July–August, pp. 39–98.

Hertz, Tom, 2005, "Rags, riches, and race: The intergenerational economic mobility of black and white families in the United States," in *Unequal Chances: Family Background and Economic Success*, Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves (eds.), Princeton, NJ: Princeton University Press.

Isaacs, Julia B., Isabel V. Sawhill, and Ron Haskins, 2008, *Getting Ahead or Losing Ground: Economic Mobility in America*, Brookings Institution and Pew Charitable Trusts, Economic Mobility Project, report, February.

Lang, Kevin, and Michael Manove, 2011, "Education and labor market discrimination," *American Economic Review*, Vol. 101, No. 4, June, pp. 1467–1496.

Mazumder, Bhashkar, 2011, "Black-white differences in intergenerational mobility in the US," Federal Reserve Bank of Chicago, working paper, No. 2011-10, November.

__________, 2008, "Upward intergenerational economic mobility in the United States," Pew Charitable Trusts, Economic Mobility Project, report, May 29.

Neal, Derek A., and William R. Johnson, 1996, "The role of premarket factors in black-white wage differences," *Journal of Political Economy*, Vol. 104, No. 5, October, pp. 860–895.

Smith, James P., and Finis R. Welch, 1989, "Black economic progress after Myrdal," *Journal of Economic Literature*, Vol. 27, No. 2, June, pp. 519–564.

Nowcasting using the Chicago Fed National Activity Index

Scott A. Brave and R. Andrew Butters

Introduction and summary

The Chicago Fed National Activity Index (CFNAI) is a monthly index of U.S. economic activity constructed from 85 data series (or indicators) classified into four groups: production and income; employment, unemployment, and hours; personal consumption and housing; and sales, orders, and inventories.[1] The index is estimated as the first principal component of the 85 data series,[2] and is essentially a weighted average of the indicators, with their individual weights representing the relative degree to which each indicator explains the overall variation among them. The CFNAI is also normalized to reflect deviations around a long-term historical rate of economic growth. As such, a zero value of the index indicates that growth in economic activity is proceeding along its long-term historical path; a negative value indicates below-average growth, while a positive value indicates above-average growth.

The CFNAI, which premiered in March 2001, was originally designed as a leading indicator for inflation (Stock and Watson, 1999; and Fisher, 2000). However, much of its current value derives from its ability to capture U.S. business cycles (that is, the periodic fluctuations in economic activity around its long-term historical trend) and nowcast[3] U.S. real gross domestic product (GDP) growth (Evans, Liu, and Pham-Kanter, 2002; and Brave and Butters, 2010). The index has been shown to align with the historical timing of U.S. recessions according to the National Bureau of Economic Research (NBER), with close to 95 percent accuracy (Berge and Jordà, 2011). Moreover, the CFNAI has the ability to signal in real time the onset and end of a recession—for instance, the index did this for the 2001 and 2007–09 recessions within one to three months of the NBER dates, with an average lead time of one year prior to the official NBER announcements (Brave and Butters, 2010). The CFNAI's success has been more mixed in terms of predicting real GDP growth, although for the 2004–09 period its performance was on par with the median current quarter forecast from the Federal Reserve Bank of Philadelphia's *Survey of Professional Forecasters* (Brave and Butters, 2010).

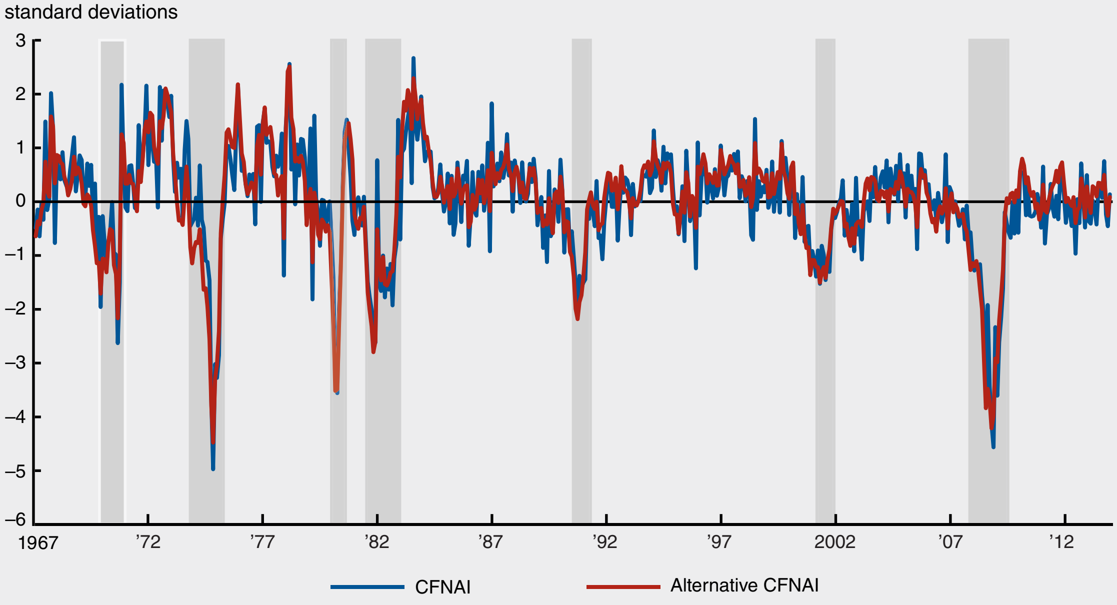
In this article, we consider an alternative version of the CFNAI that is chiefly constructed using the methodology developed in Bräuning and Koopman (2014). Their method of collapsed dynamic factor (CDF) analysis[4] offers several advantages over the CFNAI's traditional methodology—principal components analysis (PCA)—when it comes to estimating the index: first, through its incorporation of the dynamic properties

Scott A. Brave is a senior business economist in the Economic Research Department at the Federal Reserve Bank of Chicago, and R. Andrew Butters is a graduate student at the Kellogg School of Management, Northwestern University. *The authors thank Alejandro Justiniano, Dick Porter, and an anonymous referee for helpful comments and suggestions.*

ISSN 0164-0682

FIGURE 1

CFNAI versus alternative CFNAI



Notes: The figure displays the monthly Chicago Fed National Activity Index (CFNAI) and the alternative CFNAI, which is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details). Both indexes are shown in standard deviation units from their means over the period March 1967 through February 2014. Shading indicates U.S. recessions as identified by the National Bureau of Economic Research.

Source: Authors’ calculations based on data from Haver Analytics.

of the time series for the index (which PCA cannot exploit) and, second, by further disentangling common drivers of the variation in the underlying data series from idiosyncratic ones. Common drivers of the CFNAI indicators are the types of macroeconomic shocks generally associated with the business cycle, while idiosyncratic drivers include shocks typically isolated to various specific sectors of the U.S. economy, as captured in the four broad categories of the CFNAI indicators. Moreover, the methodology of Bräuning and Koopman (2014) makes it possible to directly link the CFNAI to broad economic indicators constructed at a lower frequency, such as quarterly real GDP growth.

Figure 1 plots the history of the traditional monthly CFNAI and the alternative CFNAI, which is largely based on applying the methodology of Bräuning and Koopman (2014), from March 1967 through February 2014. The shaded periods in the figure represent U.S. recessions as identified by the NBER. The alternative CFNAI shown here produces a superior in-sample fit and out-of-sample projections of current quarter real GDP growth while correlating more closely with NBER recessions than the traditional CFNAI. These improvements depend on both the way in which the correlation structure of the 85 underlying data series (at a certain point in time and across time) is taken into account in the estimation procedure and the particular way in which real GDP growth and its dynamic properties are incorporated. We establish this fact by drawing comparisons between the static factor model for the CFNAI and several dynamic factor models that include quarterly real GDP growth but differ from the Bräuning and Koopman (2014) methodology in *how* they include it.

In the process of updating the CFNAI with the Bräuning and Koopman (2014) methodology, we also learn something about the nature of the recovery from the most recent recession. Our application of CDF analysis provides us with some additional context for the uneven pattern of growth during the recovery: First, we show that there has been a moderate decline in the trend rate of real GDP growth since December 2007; and second, we note that the share of variation among the underlying data series for the CFNAI (particularly those in the personal consumption and housing category and employment, unemployment, and hours category) due to idiosyncratic drivers increased during the 2007–09 recession and the subsequent recovery. The second finding leads our estimate of the alternative CFNAI during much of the recent recovery to be higher than that of the traditional CFNAI given the weakness of these

BOX 1

Principal components analysis and factor analysis

Here, we explain the mathematics behind using PCA to construct the traditional CFNAI. Let x denote the $N \times 1$ column vector of N data series at time t. The first step is to form the $N \times T$ matrix of data vectors X_t, where each row of this matrix contains T observations normalized to have a mean of zero and a standard deviation of one.[1] The eigenvector–eigenvalue decomposition of the variance–covariance matrix $\frac{X_t X_t'}{N}$ then produces a set of time-invariant weights referenced by the 1×85 row vector w resulting from a transformation of the eigenvector associated with the largest eigenvalue of this matrix. These weights are then used to construct a weighted average of the x such that the resulting index is given by $I = wx$.

The underlying assumption about X_t necessary to produce this variance decomposition is that it admits a factor model representation. This means that it can be additively decomposed into the product of two vectors—an $N \times 1$ column vector of time-invariant factor loadings Γ and a $1 \times T$ time-varying latent factor F_t—and a normally distributed mean zero random variable ε_t with variance–covariance matrix $\sigma^2 I$:

$$X_t = \Gamma F_t + \varepsilon_t.$$

The values of F_t and Γ are then jointly estimated by maximizing $\mathrm{Tr}[\Gamma' X_t X_t' \Gamma]$ subject to a normalization constraint.[2] This linear optimization problem is solved by setting the estimator $\frac{\hat{\Gamma}'}{N}$ equal to w.[3] The estimated factor in our example, given by $\hat{F}_t = \frac{\hat{\Gamma}' X_t}{N}$, corresponds to the CFNAI. As such, it is the principal component common to all $N = 85$ indicators that explains the largest amount of variation among them.

[1]Underlying the normalization of the data is the concept of stationarity, or in this case the first and second moment restrictions that the mean and variance of each indicator do not vary over time. Each data series first receives a transformation to make it stationary prior to its normalization. A list of transformations can be found at www.chicagofed.org/digital_assets/publications/cfnai/background/cfnai_indicators_list.pdf.

[2]The normalization constraints most commonly used with one factor are $\frac{\Gamma'\Gamma}{N} = 1$ and $\frac{F_t' F_t}{T} = 1$.

[3]See Stock and Watson (2002) for more details on the connection between PCA and factor analysis. Also, note that identification here is achieved only up to the scale provided by the normalization constraint on the factor loadings. To be able to interpret the CFNAI, we use the normalization $\frac{F_t' F_t}{T} = 1$.

data series, although the first finding somewhat offsets the impact on U.S. real GDP growth of this upward revision in our assessment of economic activity.

In the next section, we detail the traditional and several alternative methods of constructing the CFNAI. Then, we further explain the implications of our proposed update to the CFNAI using the Bräuning and Koopman (2014) framework. Next, we describe what is driving the differences in timing of U.S. recessions across the Chicago Fed National Activity Index's three-month moving average (CFNAI-MA3) and the alternative CFNAI's three-month moving average (alternative CFNAI-MA3). Afterward, we show how the alternative CFNAI-MA3 can be used to nowcast annualized quarterly real GDP growth more accurately than the traditional CFNAI-MA3 and several other dynamic-factor-based indexes. Finally, we present our conclusions and comment on what they may imply for U.S. economic activity over the near term.

Traditional and alternative methods of constructing the CFNAI

The traditional method of constructing the CFNAI—principal components analysis—proceeds by means of an eigenvector–eigenvalue decomposition of the variance–covariance matrix of the 85 underlying data series. This *static* factor model description of the data, detailed in box 1, produces a principal component for each eigenvalue of the variance–covariance matrix. The eigenvector associated with the largest eigenvalue of the matrix constitutes the weights applied to the data series that are used to construct the first principal component, or what we call the CFNAI. Stock and Watson (2002) show that this method of constructing the CFNAI is capable of producing a consistent estimate of the underlying static factor model of the data as the number of data series and the number of time periods become large.

The CFNAI is estimated monthly and released near the middle of each month with a history from March 1967 through the month preceding that of the release date. The lag of approximately one month between the last month of the index and the release date is necessary because of limitations on data availability. In addition to this one-month production lag, many of the data series themselves are only available at a further one- to two-month lag. The limited availability of data at the time of estimation results in a variance–covariance matrix of less than full rank, thereby making principal components analysis infeasible. To circumvent this issue, we forecast each incomplete data series separately up to the month in which the index is produced according to individual autoregressive processes with five lags before the index is constructed.

This method of completing the panel of indicators in order to construct a first principal component is very flexible but not unique or even necessary. Stock and Watson (2002) also demonstrate how to produce an index estimate when data are missing with the same desirable statistical properties as PCA. Their methodology relies explicitly on the incomplete data methods of the expectation–maximization (EM) algorithm made popular by Watson and Engle (1983); and although it does not take into account the serial correlation properties of each data series like the current CFNAI procedure for inferring missing data or the dynamic properties of the index itself, it does account for the data's underlying factor representation to both estimate the index and impute missing values.[5]

The Stock and Watson (2002) EM algorithm uses the information from the complete, or "balanced," panel of indicators to make the best possible prediction of the incomplete, or "unbalanced," panel of indicators. When applied to the construction of the CFNAI, it begins by performing PCA on the subset of data series that are available in all time periods. Missing values are then predicted based upon linear regressions of each of the 85 data series on the first principal component. Finally, PCA is repeated on the balanced panel of data, which combines the observed and predicted data. This process continues until the difference in the sum of the squared prediction errors between iterations reaches a desired level of convergence.

Since the inception of the CFNAI in March 2001, several alternative methods of constructing economic activity indexes that build on PCA have been proposed. Each of these is also an example of factor analytic methods; the differences across the methodologies mainly depend on how variation due to common drivers versus idiosyncratic ones is decomposed across data series and time periods. Here, we briefly describe a few of these alternative methodologies, contrasting them with the traditional methodology used for the CFNAI and each other before we explain the collapsed dynamic factor model of Bräuning and Koopman (2014). Boxes 2 and 3 present many of the technical details of these methods that are omitted in the discussion that follows.

Giannone, Reichlin, and Small (2008) extended the static representation of the factor model of Stock and Watson (2002) into a *dynamic* factor model by incorporating both information from the cross section of data series (at each point in time) and information on data series across time into the process of estimating the index and imputing missing values. Doz, Giannone, and Reichlin (2012) then subsequently provided an alternative EM algorithm with which to estimate the dynamic factor model. In the first step, PCA is performed up to the point in time for which all data series are available. The first principal component from this static factor model is then used to obtain the initial parameter values for the dynamic factor model shown in box 2. The estimation of the CFNAI in this case proceeds by means of the Kalman filter and smoother equations applied to this model. The resulting index is then used to reestimate parameter values, and the process is repeated until convergence of the model's log-likelihood is achieved as shown in box 2.

In the methodology of Doz, Giannone, and Reichlin (2012), data series that are unavailable each month are ignored for inferring the value of the index, but are forecasted using information on the dynamic properties of the index via the Kalman filter. Furthermore, unlike PCA, the idiosyncratic error structure of the data can be relaxed to accommodate unequal variances across unobserved idiosyncratic drivers of the data series (that is, heteroskedasticity). These modifications of the underlying factor model for the CFNAI are not costless, however, as they come at the price of estimating a much larger number of parameters. Besides the obvious increase in complexity and in the time and computing power necessary to estimate and construct the index using dynamic factor methods versus static factor methods, other potential drawbacks from this richer class of factor models include the additional uncertainty introduced when using the index to make out-of-sample projections of inflation and economic growth as in Fisher (2000) and Brave and Butters (2010).

Collapsed dynamic factor analysis as presented in Jungbacker, Koopman, and van der Wel (2011) and explained in box 3 minimizes these costs by transforming the static portion of the dynamic factor model in such a way as to significantly reduce the number of estimated parameters needed to run the Kalman filter and compute projections of auxiliary variables of interest, such as real GDP growth. We follow their methodology in order to incorporate serial correlation within data series (idiosyncratic autocorrelation) in addition to heteroskedasticity into the static factor model. All of the alternative methods for constructing the CFNAI that we have presented thus far, however, remain sensitive to the use of PCA as a starting point for estimation.[6] If the PCA estimate of the index is biased even after accounting for its dynamics, none of the dynamic factor models considered here is guaranteed to produce an unbiased estimate of the index.

Bräuning and Koopman (2014) provide an alternative transformation of the static factor model that does

BOX 2

Dynamic factor analysis

PCA and traditional factor analysis are static estimation methods in that they do not incorporate information from both the cross section of data series and the information from across time. Dynamic factor analysis instead makes use of variation in both forms. To do so, it relies on signal extraction methods, such as the Kalman filter, applied to a system of equations relating the latent factor, or the CFNAI as in the example in box 1 (p. 21), to both the cross section of data series at each point in time (a "measurement" or "observation" equation) and the dynamic factors that drive its fluctuations over time (a "state" equation).

Mathematically, this involves specifying the following state-space representation:

$$X_t = \Gamma F_t + \varepsilon_t,$$

$$F_t = AF_{t-1} + \upsilon_t,$$

where F_t is the $1 \times T$ latent factor capturing a time-varying common source of variation in the $85 \times T$ matrix of indicators X_t; Γ is the 85×1 loadings onto the factor; and A is the transition matrix describing the evolution of the latent factor over time. We write the A parameter of the model assuming a first-order autoregressive process (AR(1)) for F_t, which can be generalized to an arbitrary number of lags, p.[1] The static factor model representation of the CFNAI described in box 1 thus forms the measurement equation of the state-space representation of the dynamic factor model. Adding dynamics of some finite order to F_t yields its state equation.

Both ε_t and υ_t are assumed to be independently normally distributed mean zero random variables. We follow the dynamic factor model of Doz, Giannone, and Reichlin (2012) and assume that $\mathrm{Var}(\varepsilon_t) = H$ (an 85×85 diagonal matrix) and $\mathrm{Var}(\upsilon_t) = 1$.[2] The signal extraction methods of the Kalman filter and smoother are capable of estimating such a model given the coefficient matrices of the measurement and state equations, that is, Γ and A, and the idiosyncratic error variances along the diagonal of H. All of these parameters can be consistently estimated from linear regressions involving X_t and the smoothed or PCA estimate of F_t as demonstrated in Giannone, Reichlin, and Small (2008).

With the model in state-space form and initial estimates of the system matrices, the expectation–maximization (EM) algorithm outlined by Shumway and Stoffer (1982) can be used to estimate the latent factor F_t. At each iteration of the algorithm, one pass of the data through the Kalman filter and smoother is made followed by reestimating the system matrices.[3] The log-likelihood that results is nondecreasing, and convergence is governed by its stability.[4] This iterative estimation process combines the efficiency of likelihood-based estimation of the latent factor with the consistency of ordinary least squares (OLS) parameter estimates.

To see the relationship between the static and dynamic factor models, consider the case where the transition matrix of the state equation, A, is the zero matrix. That is, nullify the impact of dynamics for the latent factor. Notice that if we specify the variance–covariance matrix of the measurement equation's error term is proportional to the identity matrix (and based on the description of PCA discussed in box 1), we end up with an estimate of the latent factor that is proportional to the first principal component. For this reason, our traditional methodology for the CFNAI can be considered a special case of the dynamic factor model with a zero transition matrix and a homoskedastic idiosyncratic error structure (that is, the assumption of equal variances across unobserved idiosyncratic drivers of the underlying data series).

[1]We choose p depending on the model being estimated, but all models use either three or four lags.

[2]The latter restriction acts to set the scale of the dynamic factor model just as the normalization on the scale of the factor loadings used in PCA does for the static factor model.

[3]In addition, a small alteration in the least-squares step is required to account for the fact that the unobserved components of the model must first be estimated. See Durbin and Koopman (2012) for further details.

[4]Our stability criterion where k references iteration is as follows: $|\log \mathrm{L}(k) - \log \mathrm{L}(k-1)/((\backslash\log \mathrm{L}(k) + \backslash\log \mathrm{L}(k-1))/2)| < 10^{-6}$.

not assume that PCA produces an unbiased estimate of the index. In their framework, each data series' factor loadings are fixed at their PCA values. Then, by treating the PCA estimate of the index as a noisy indicator of the "true" measure, their method reoptimizes the index such that it explains the largest percentage of the variation in the PCA estimate of the index that is consistent with both its own estimated dynamics and that of an auxiliary "target" variable. The target variable is described as a more comprehensive but perhaps less frequently available indicator of the information set over which the other data series span. The target variable is also unique in that it alone follows its own estimated dynamics and loads directly on both current and past values of the index, whereas in our application the dynamics of the index do not depend on the target variable. In our application and theirs, real GDP growth is used as the target variable.

Our use of the Bräuning and Koopman (2014) methodology is motivated by several persistent criticisms

BOX 3

Collapsed dynamic factor analysis

Collapsed dynamic factor analysis reflects its name. It begins by applying a transformation to the measurement equation of the dynamic factor model's state-space representation in order to collapse its size to match the typically smaller size of the state equation. In the context of the Jungbacker, Koopman, and van der Wel (2011) model applied to the CFNAI, this amounts to premultiplying the 85 × T matrix of indicators X_t by the transformation $A_L = (\Gamma'\Omega^{-1}\Gamma)^{-1}\Gamma'\Omega^{-1}$, where Ω is the variance–covariance matrix of ε_t:

$$X_t^L = F_t + u_t.$$

The transformed measurement equation, shown here, then relates a scalar, $X_t^L = A_L X_t$, with a unit factor loading to the latent factor, F_t, and a mean zero normally distributed random scalar u_t with variance $H = (\Gamma'\Omega^{-1}\Gamma)^{-1}$. The state equation is unaltered from the example in box 2 (p. 23).

Notice that the transformation here when applied to X_t takes the familiar form of the generalized least squares (GLS) solution for the latent factor F_t with Ω as the weight matrix. Bräuning and Koopman (2014) suggest the use of an alternative transformation. In their example, $A_L = \frac{\hat{\Gamma}'}{N}$, where $\hat{\Gamma}'$ is the PCA estimate of the factor loadings of the static factor model, $X_t^L = \frac{\hat{\Gamma}' X_t}{N}$ is the traditional CFNAI as shown in box 1 (p. 21), and $u_t = \frac{\hat{\Gamma}'\varepsilon_t}{N}$. Furthermore, H is not assumed to be a predetermined function of the dynamic factor model's factor loadings and the variance–covariance matrix of its idiosyncratic errors. It is instead estimated as an additional parameter. The estimation of H is made possible by the inclusion of an additional measurement equation containing a "target" variable, which is real GDP growth in their example and ours.

The random scalar u_t in this context has the interpretation of a "measurement error" between the PCA estimate of the CFNAI and its dynamic factor counterpart. Notice that it is also a weighted average of the idiosyncratic disturbances of the static factor model, with the weights corresponding to the PCA factor loadings. The implicit assumption maintained by Bräuning and Koopman (2014) to derive their transformation is that $\frac{\hat{\Gamma}'\hat{\Gamma}}{N} \cong 1$. Deviations from this assumption will produce some approximation error as well in u_t.

We modify the Bräuning and Koopman (2014) methodology for our purposes by applying the transformation to an alternative representation of the indicators, $\tilde{X}_t = X_t - \rho X_{t-1}$. This modification allows us to draw finer comparisons with the collapsed dynamic factor model of Jungbacker, Koopman, and van der Wel (2011), which also allows for heteroskedasticity and serial correlation in the idiosyncratic errors ε_t but assumes PCA produces an unbiased estimate of the latent factor. To make this modification operative, we first estimate the collapsed dynamic factor model of Jungbacker, Koopman, and van der Wel (2011) to obtain estimates of the ρ vector and construct $\tilde{X}_t$.[1] We then apply PCA to the covariance matrix of $\tilde{X}_t$ to obtain $A_L = \frac{\hat{\Gamma}'}{N}$ and proceed as described earlier in this box.[2]

Our application of the Bräuning and Koopman (2014) methodology also requires an additional measurement equation relating quarterly real GDP growth, Y_t, to its own lagged value, Y_{t-3}; current and past values of the three-month moving average of the CFNAI, F_t^3; and a time-varying intercept, T_t^3. Real GDP growth in this framework acts to "clean" the PCA estimate of the three-month moving average of the monthly index by apportioning it in each quarter into a fragment that is correlated with quarterly real GDP growth, $\gamma_0 F_t^3 + \Sigma_{k=1}^3 \gamma_k F_{t-k}^3$, and a fragment that is not, $T_t^3 + \delta Y_{t-3} + v_t$, based on the regression coefficients, γ_k and δ. The mean zero normally distributed random variables u_t and v_t are assumed to be independent. Box 4 provides more details on this particular nowcasting specification:

$$Y_t = T_t^3 + \delta Y_{t-3} + \gamma_0 F_t^3 + \sum_{k=1}^{3} \gamma_k F_{t-k}^3 + v_t.$$

This errors-in-variables framework is estimated by Bräuning and Koopman (2014) by full maximum likelihood techniques. Here, to maintain consistency with the way the other dynamic and collapsed dynamic factor models are estimated, we instead use a variant of the EM algorithm described in box 2 to estimate the transformed state-space representation. In order to use the Jungbacker, Koopman, and van der Wel (2011) estimate of the smoothed latent factor in the first step, this process requires a restricted least-squares regression of the PCA estimate of the factor on the smoothed latent factor and an additional linear regression for the target variable equation. Additional details on the estimation process can be found in box 4.

[1]The maintained assumption in this exercise in order for our estimate of ρ to be unbiased is that $\mathrm{Cov}(X_{t-1}, \xi_t) = 0$, where ξ_t is a composite error term comprising ε_t and the contemporaneous measurement error in the estimated factor.

[2]Because the indicators have already been demeaned and standardized, they are measured in common units. Thus, obtaining principal components from the covariance matrix instead of the correlation matrix of $\tilde{X}_t$ allows us to incorporate unequal variances across the indicators.

of perceived bias in the CFNAI. One source of bias in the CFNAI could stem from not including enough variables to span the space of U.S. economic activity—for instance, by omitting international trade or government spending indicators (which inform real GDP growth) as the CFNAI currently does. Another source of bias in the CFNAI could be due to a preponderance of data confined to one or more sectors of the economy—for instance, the potential overweighting of manufacturing data series that dominate the production and income category of indicators and the CFNAI. Yet another source of bias in the CFNAI could result from the omission of any additional common components in the CFNAI data series in the estimation of the dynamic factor model. Here, we consider only the likelihood of the first two potential sources of bias, but note that our results remain sensitive to the possibility of the last one. See box 3 for more details on how the Bräuning and Koopman (2014) methodology helps to correct for these potential sources of bias in the CFNAI.

Implications of the update for the CFNAI

The estimation of the dynamic factor models for the CFNAI requires only slight modifications to existing methods as shown in box 3.[7] To be able to compare indexes based on alternative methodologies, we include real GDP growth as an additional indicator for each of the dynamic factor alternatives to the CFNAI's traditional methodology discussed previously. This way we can highlight the joint role played by including real GDP growth along with the dynamic factor elements discussed in the previous section. Moreover, to capture the role played by allowing for dynamics in the estimation process instead of relaxing various PCA restrictions, we use a variant of the Giannone, Reichlin, and Small (2008) methodology. In this case, the factor model for the CFNAI is estimated using the Doz, Giannone, and Reichlin (2012) EM algorithm, preserving the PCA restrictions on the idiosyncratic error structure of the data but allowing for a dynamic process of the index to be estimated.

Another benefit of the alternative estimation frameworks presented in the previous section is that, following Bräuning and Koopman (2014), it becomes feasible to decompose real GDP growth into its trend and cyclical components.[8] Based on our past work (Brave and Butters, 2010, 2013), this ability to decompose real GDP growth has turned out to be vital to capturing changes in average real GDP growth over long periods. Given this finding, we developed a specification that allows for a time-varying intercept in the equation for quarterly real GDP growth to capture changes over time in its trend rate of growth (see box 4). To capture cyclical movements, we follow Brave and Butters (2010) in using one lag of quarterly real GDP growth in addition to current and past values of the three-month moving average of the monthly index.

Figure 2 plots in separate panels the difference between the CFNAI and each of the four dynamic-factor-based indexes. Simply adding dynamic elements to the static factor model, as well as quarterly real GDP growth in the construction of the index, produces small differences from the traditional CFNAI. This can be seen in the difference between the CFNAI and the first dynamic-factor-based index (labeled DF in panel A of figure 2). Further relaxing the PCA restrictions on the idiosyncratic error structure of the data has a more pronounced effect; this is apparent in the difference between the CFNAI and the dynamic-factor-based index with heteroskedastic errors (labeled DF-HC in panel B of figure 2) and between the CFNAI and the dynamic-factor-based index with heteroskedastic and serially correlated errors (labeled DF-HAC in panel C). However, the difference from the traditional CFNAI is most prominent for the dynamic-factor-based index based on the methodology of Bräuning and Koopman (2014) (labeled CDF-HAC in panel D of figure 2)—which we refer to as the alternative CFNAI in figure 1 (p. 20). These results are detailed further in table 1 (p. 28), which displays the cumulative effect on the explained variance of the 85 underlying data series for the traditional CFNAI from altering the various assumptions underlying its static factor model. Each successive addition to the static factor model for the CFNAI—from dynamics and real GDP growth (first row, second column) to heteroskedastic errors (first row, third column) and serially correlated idiosyncratic errors (first row, fourth column)—reduces the explained variance of the 85 underlying data series by the index, but none more so than the Bräuning and Koopman (2014) methodology (first row, fifth column), which corrects for bias arising from the use of PCA. It is important to note here that the reductions in explained variance do not reflect a failure of the dynamic factor model to account for variation among these data series at a certain point in time or within them across time. Instead, such reductions reflect the fact that more of the variation in these series is estimated to arise from idiosyncratic drivers (including potential bias due to the use of PCA) rather than common ones. The alternative CFNAI (first row, fifth column) explains only 20 percent of the total variance of the 85 data series—a reduction of almost one-third of the total variance explained by the traditional CFNAI (first row, first column) and a reduction of almost one-fourth of the total variance explained by its closest counterpart, the DF-HAC index (first row, fourth column).

BOX 4

The model for nowcasting real GDP growth

Our dynamic factor model for the CFNAI is given by the system of equations in box 2 (p. 23), and is repeated here for convenience:

$$X_t = \Gamma F_t + \varepsilon_t ,$$
$$F_t = AF_{t-1} + \upsilon_t .$$

To obtain the collapsed dynamic factor models discussed in the text, we substitute the measurement equations described in box 3 (p. 24) for the first equation here.

The variant of this system based on Giannone, Reichlin, and Small (2008) parameterizes the variance–covariance matrix of ε_t, or H, as $\sigma^2 I$, in accordance with the description of PCA in box 1 (p. 21). The variant based on Doz, Giannone, and Reichlin (2012) instead assumes a heteroskedastic H with diagonal elements equal to σ_i^2. In addition to allowing for heteroskedasticity, the variant based on Jungbacker, Koopman, and van der Wel (2011) allows for idiosyncratic serial correlation up to the first order, where we choose the degree of serial correlation for each of the 85 data series prior to estimating according to the Bayesian information criterion. The CDF variant referenced in the text estimates H as a scalar parameter according to Bräuning and Koopman (2014).

We append to this model a nowcasting equation relating annualized quarterly real GDP growth, Y_t, in each time period to its own lagged value, Y_{t-3}; current and past values of the three-month moving average of the latent factor, F_t^3; and a time-varying intercept, T_t^3. We only observe Y_t in the third month of each quarter, so that this equation strictly relates each quarterly realization of real GDP growth to only the corresponding end-of-quarter value of T_t^3:

$$Y_t = T_t^3 + \delta Y_{t-3} + \gamma_0 F_t^3 + \sum_{k=1}^{3} \gamma_k F_{t-k}^3 + v_t .$$

To be able to estimate the model, we must first specify a dynamic process for the latent time-varying intercept, T_t^3, by adding a second state equation to the model. We assume that it is the quarterly average of a monthly process T_t that follows a random walk with drift parameter α:

$$T_t = \alpha + T_{t-1} + \eta_t .$$

As such, T_t^3 represents the time-varying mean of quarterly real GDP growth conditional on the previous quarter's value of real GDP growth Y_{t-3} and current and past values of F_t^3, and can be interpreted as trend real GDP growth. Furthermore, we assume that v_t and η_t are mean zero normally distributed random variables with variances V and W, respectively, that are uncorrelated with each other, ε_t, and υ_t.

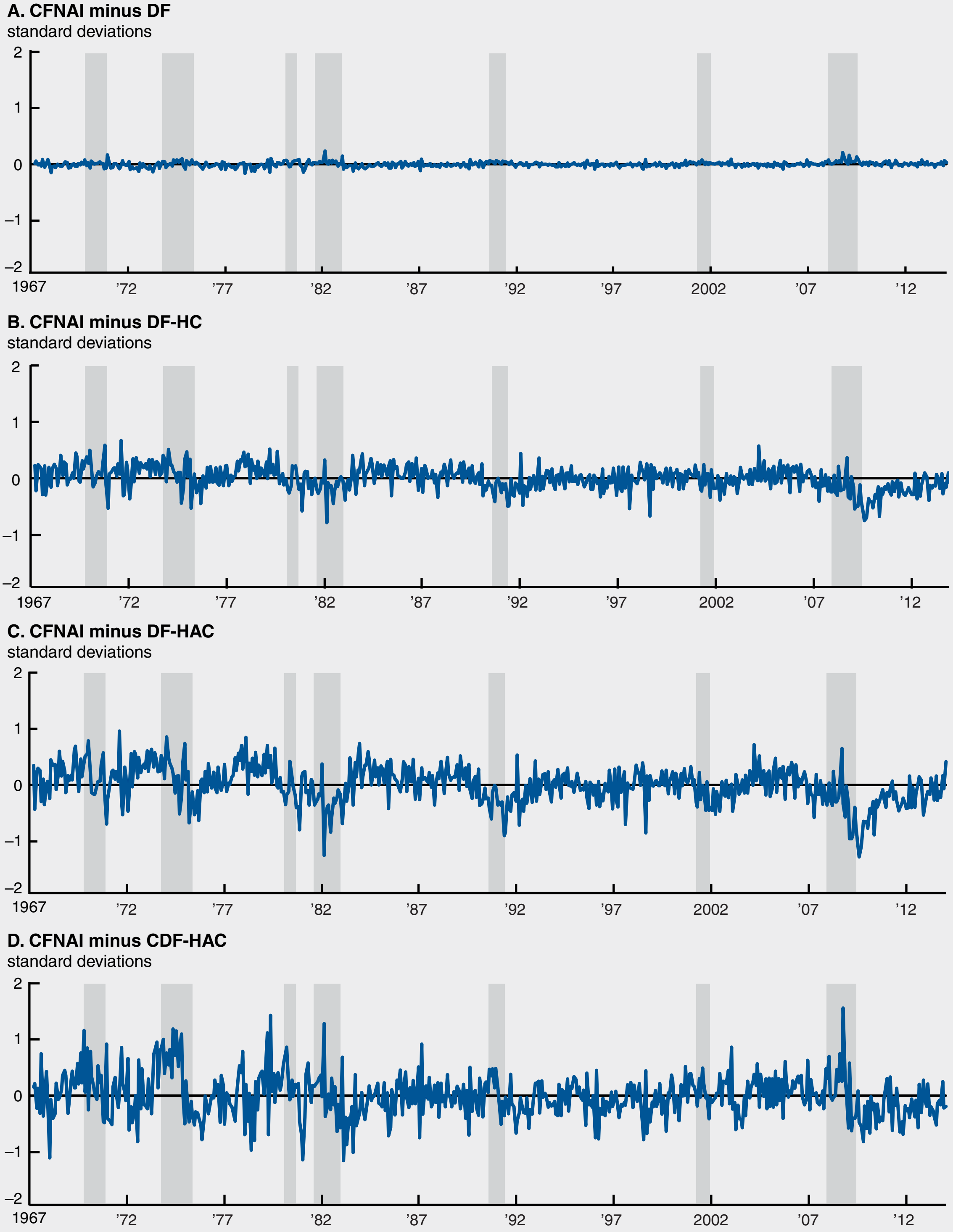
This particular specification of the nowcasting equation expands on Brave and Butters (2010), in which we used the CFNAI to nowcast real GDP growth, and is largely taken from the follow-up discussion in Brave and Butters (2013). It is based on a decomposition of trend and cyclical components for real GDP growth as in Bräuning and Koopman (2014), where the cyclical dynamics of real GDP growth are assumed to be captured by lagged real GDP growth and current and past values of the three-month moving average of the latent factor. However, it also represents a departure from the specification considered by Bräuning and Koopman (2014), which uses a different method of aggregation to relate real GDP growth to the monthly latent factor, includes additional lags of real GDP growth, and does not include a time-varying intercept.

Our model is estimated using a variant of the EM algorithms described in boxes 2 and 3. The use of the Kalman filter requires that we specify initial values for the mean and variance of F_t and T_t. Here, we use the exact initialization procedure described in Harvey (1989) for F_t, as well as a diffuse initialization for T_t by assuming that its initial mean value is the estimated constant in the presample regression of annualized quarterly real GDP growth on a constant in the 20 quarters prior to our sample beginning in March 1967 and setting its initial variance to the variance of this estimate. From the in-sample regression of annualized quarterly real GDP growth on a constant, one lag of itself, and current and previous values of the CFNAI-MA3, we then obtain our initial parameter estimates of δ, γ, and V. Initializing α at zero, we then obtain our initial estimate of W according to the median unbiased estimation procedure described in Stock and Watson (1998) applied to a local-level unobserved components model for quarterly real GDP growth. At subsequent iterations, α and W are then reestimated by restricted linear regression using our estimate of T_t.

Overall, correcting for bias is of greater importance than any other modification in explaining the differences between the traditional CFNAI and the alternative CFNAI according to the results in table 1. However, the other modifications to the underlying static factor model for the CFNAI reflected in the table are also worth highlighting. For instance, the various dynamic-factor-based indexes exhibit very different shares of explained variance by the index across the four broad categories of indicators. Allowing for heteroskedastic

FIGURE 2

Differences between the CFNAI and four dynamic-factor-based indexes



Notes: The figure displays the differences between the monthly Chicago Fed National Activity Index (CFNAI) and each of the four dynamic-factor-based indexes—DF, DF-HC, DF-HAC, and CDF-HAC—derived from methodologies based on Giannone, Reichlin, and Small (2008), Doz, Giannone, and Reichlin (2012), Jungbacker, Koopman, and van der Wel (2011), and Bräuning and Koopman (2014), respectively (see the text for further details). Each index was standardized (that is, transformed to have a zero mean and a standard deviation of one) prior to calculating the differences so that the displayed units are standard deviations from the respective index means. The differences are plotted over the period March 1967 through February 2014. Shading indicates U.S. recessions as identified by the National Bureau of Economic Research.
Source: Authors’ calculations based on data from Haver Analytics.

TABLE 1

Fraction of data variance explained by the index

	CFNAI	DF	DF-HC	DF-HAC	CDF-HAC
Total	0.29	0.28	0.27	0.26	0.20
Production and income	0.39	0.38	0.46	0.50	0.43
Employment, unemployment, and hours	0.36	0.37	0.33	0.29	0.32
Personal consumption and housing	0.08	0.08	0.05	0.03	0.04
Sales, orders, and inventories	0.17	0.17	0.16	0.18	0.21

Notes: The table displays the fraction of the overall variance of the 85 underlying indicators in the Chicago Fed National Activity Index (CFNAI) that is explained by the CFNAI and each of the four dynamic-factor-based indexes over the period March 1967 through February 2014. In addition, it decomposes this fraction into the share explained by each of the four broad categories of indicators listed here. The four dynamic-factor-based indexes—DF, DF-HC, DF-HAC, and CDF-HAC—are derived from methodologies based on Giannone, Reichlin, and Small (2008), Doz, Giannone, and Reichlin (2012), Jungbacker, Koopman, and van der Wel (2011), and Bräuning and Koopman (2014), respectively (see the text for further details).
Source: Authors' calculations based on data from Haver Analytics.

errors shifts explained variance toward the production and income category of indicators and away from the other three categories (see second through fifth rows, differences between second and third columns). Additionally allowing for idiosyncratic autocorrelation has a similar effect but also boosts the share of explained variance due to the sales, orders, and inventories category (see second through fifth rows, differences between third and fourth columns). The employment, unemployment, and hours category and personal consumption and housing category are particularly affected by the modifications to the idiosyncratic error structure of the static factor model for the CFNAI.[9] For these reasons (and as explained in box 3, p. 24), we deviate slightly from the Bräuning and Koopman (2014) model by continuing to account for both heteroskedastic and serially correlated errors in the CDF-HAC index.

Additionally, allowing (and correcting) for bias from using PCA in the estimation of the CFNAI, as shown in the fifth column of table 1, serves to reapportion the explained variance shares slightly more equally among the remaining three categories at the expense of the production and income category of indicators. In fact, much of the bias in the CFNAI that we estimate can be traced back to the contribution of the production and income category of indicators. Hence, the concern over potential overweighting of manufacturing data sources that dominate this category of indicators appears to be valid. The end result is an index (that is, the CDF-HAC index) that puts slightly more weight on the sales, orders, and inventories and production and income categories than the traditional CFNAI does (despite the correction for bias arising from the latter category) and less weight on the personal consumption and housing and employment, unemployment, and hours categories. Furthermore, we should point out that although the difference in the personal consumption and housing category's share of the fraction of data variance explained by the CFNAI and the alternative CFNAI (the CDF-HAC index) may at first seem small, its economic significance is anything but small given the outsized contribution of this category to the weakness in economic activity during the recent recession and subsequent recovery. In fact, we find that a sizable portion of the upward revision seen in the alternative CFNAI during the recovery can be traced back to this result, as we discuss in the next section.

Capturing business cycles

One of the CFNAI's key successes has been its use as an indicator of U.S. business cycles. Traditionally, the three-month moving average of the index—the CFNAI-MA3—has been used for this purpose in the past on account of the volatile nature of the monthly CFNAI. We follow this precedent here, but note that one clear benefit of the Bräuning and Koopman (2014) methodology is that it mitigates to some degree the concern about the volatility of the monthly index. Using the nonparametric method developed in Berge and Jordà (2011), we can quantify the accuracy of both the CFNAI-MA3 and the three-month moving average of the alternative CFNAI in capturing U.S. expansions and recessions as defined by the NBER.[10] The receiver operating characteristic (ROC) analysis framework that Berge and Jordà describe produces a simple summary statistic in this regard (the area under the receiver operating characteristic curve, or AUROC). We briefly explain how we use this method next, while technical details for our ROC analysis can be found in box 5.

Our use of ROC analysis can be explained graphically by a histogram, as shown in figure 3. This figure plots the relative frequency of every observed value of the alternative CFNAI-MA3 separately for values that occur during NBER recessions and expansions. One

BOX 5

Receiver operating characteristics analysis

ROC analysis applied to the CFNAI and its dynamic-factor-based alternatives requires that we categorize each observation of an index as falling within a recession or expansion. Following the dating conventions for U.S. business cycles of the NBER, we then need to construct these conditional probabilities:

$$TP(c) = P[I_t \geq c | S_t = 1],$$

$$FP(c) = P[I_t \geq c | S_t = 0],$$

with $S_t \in \{0, 1\}$ indicating recessions and expansions, respectively. $TP(c)$ is typically referred to as the true positive rate, and $FP(c)$ is known as the false positive rate for an index I_t and particular observed value c. The relationship between the two is described by the ROC curve. With the Cartesian convention, this curve is given by

$$\{ROC(r), r\}_{r=0}^{1},$$

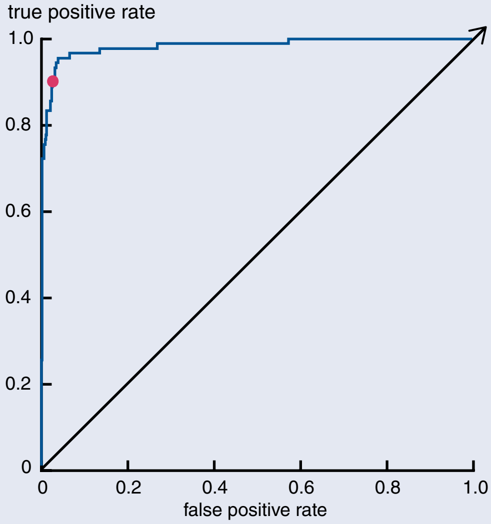
where $ROC(r) = TP(c)$ and $r = FP(c)$. In what follows, we describe how to construct the ROC curve.

Using the data in figure 4 (p. 32), we find the fraction of observations that fall outside and inside the shaded regions denoting U.S. recessions according to the NBER for the alternative CFNAI-MA3. These fractions are the *unconditional* probabilities associated with expansions and recessions. To obtain *conditional* probabilities, we use the following algorithm: For each value between the minimum and maximum observations of an index, we find the fraction of observations where that value and all subsequently higher values fall outside the shaded regions. We then do the same to find the fraction of observations that fall inside the shaded regions. These two statistics are equivalent to the *true* and *false positive* rates for separating expansions from recessions defined previously. By plotting the true and false positive rates against each other for every historical value of an index, we produce a nonparametric estimate of its ROC curve.

Berge and Jordà (2011) show that by calculating the AUROC we arrive at an estimate of the ability of the index to delineate recessions from expansions. As the area under the curve approaches 1, the more predictive it is of U.S. expansions and recessions; its statistical significance is judged relative to the area under the line from the origin extending at a 45-degree angle (see the next paragraph for more details).[1] It is also possible to compare the area under two different curves to distinguish the statistical significance of differences in predictive ability. This technique is commonly used in the medical statistics literature to evaluate the ability of a procedure or medical test to distinguish patients afflicted with a condition from those who are not.[2]

FIGURE B1

The ROC curve for the alternative CFNAI-MA3



Notes: The solid blue line is the receiver operating characteristic (ROC) curve for the three-month moving average of the alternative Chicago Fed National Activity Index (alternative CFNAI-MA3), which is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details), and U.S. expansions and recessions as identified by the National Bureau of Economic Research over the period March 1967 through February 2014. The solid black line is a 45-degree line from the origin, with a slope equal to 1 and an area under the line equal to 0.5. The red dot on the ROC curve corresponds with the recession prediction threshold explained in the text of box 5.

Source: Authors' calculations based on data from Haver Analytics.

Figure B1 displays the ROC curve for the alternative CFNAI-MA3 along with a line from the origin at a 45-degree angle. By construction, this line has an AUROC equal to 0.5. The more the ROC curve deviates in total above this 45-degree line, the higher an index's AUROC will be. In addition, for an index's AUROC to exceed 0.5, it must have a slope greater than 1 at some point on the ROC curve such that, for a given increase in the true positive rate, the associated increase in the false positive rate is smaller. The red dot on the curve marks the point at which it is no longer possible to increase the true positive rate without producing more false positives than are consistent with the observed relative frequency of expansions and recessions.

BOX 5 (CONTINUED)

Receiver operating characteristics analysis

Baker and Kramer (2007) show that the point on the curve denoted in figure B1 by the red dot meets the decision-theoretic criteria for a threshold rule, c^*, that equally penalizes type I (false positive) and type II (false negative) classification errors for recessions and expansions. To see this, consider the following utility function:

$$U = U_{11}ROC(r)\pi + U_{01}\big(1 - ROC(r)\big)\pi + U_{10}r(1-\pi) + U_{00}(1-r)\,(1-\pi),$$

where U_{ij} is the utility (or disutility) associated with the prediction i given that the true state of the business cycle, S_t, is j, with $\{i, j\} \in \{0, 1\}$ and where π is the unconditional probability of an expansion. Utility maximization implies the following first-order condition determining c^*:

$$\frac{\partial ROC}{\partial r} = \frac{U_{00} - U_{10}}{U_{11} - U_{01}} \frac{1-\pi}{\pi}$$

If we set the leading ratio of utilities to 1, this threshold equates the slope of the ROC curve to the ratio of the unconditional probabilities of expansion and recession. In doing so, one is essentially equally weighting the net benefit of making a type I error versus a type II error relative to correctly predicting the true state of the business cycle.

[1]The procedure for evaluating statistical significance is described in DeLong, DeLong, and Clarke-Pearson (1988).

[2]See Brave and Butters (2012a, 2012b) for further examples using this approach to predict financial crises.

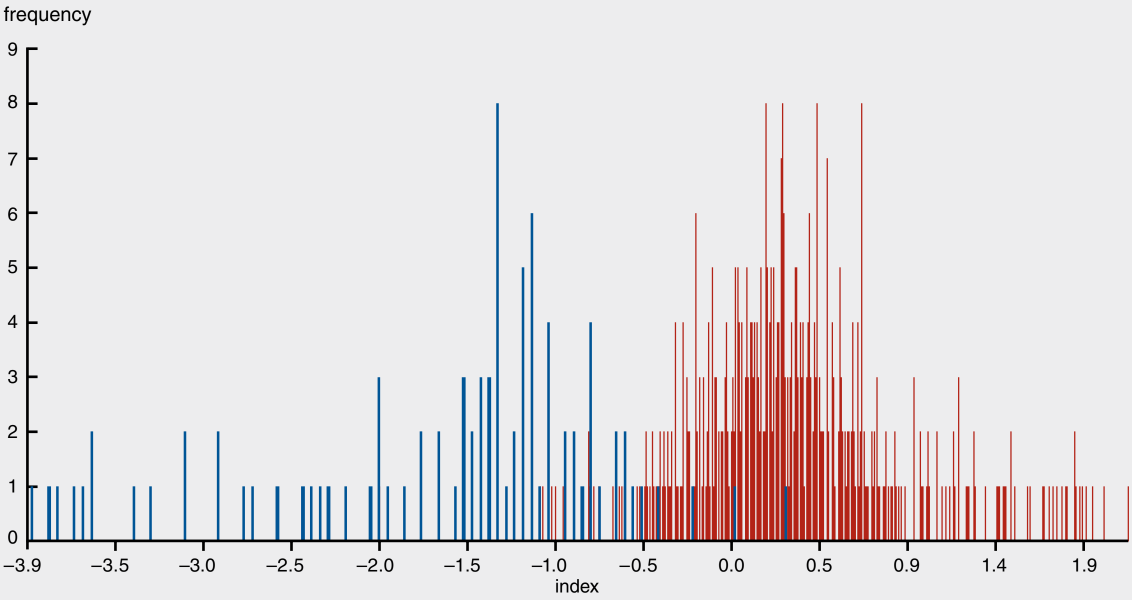
can see from figure 3 that the alternative CFNAI-MA3 is in fact quite accurate at separating recessions from expansions, as the empirical distributions seldom overlap. The AUROC statistic measures the degree of separation of the two distributions, such that the more accurate an index is at distinguishing expansions from recessions, the higher its AUROC value will be. As noted in box 5, it is even possible to compare two AUROC values to assess whether or not their differences are statistically significant. The CFNAI-MA3 has 94 percent accuracy in describing NBER expansions and recessions, so surpassing its level of accuracy in this respect is a tall task for any of the dynamic-factor-based indexes to achieve; however, one—the three-month moving average of the alternative CFNAI (CDF-HAC index)—does in fact surpass the CFNAI-MA3's accuracy at the 95 percent confidence level, with an AUROC of 98 percent. None of the other three-month moving averages of the dynamic-factor-based indexes we considered were able to produce a statistically significant improvement in AUROC compared with the CFNAI-MA3, as shown in the first column of table 2. Yet, it was true for the alternative CFNAI regardless of whether or not we smoothed through some of the monthly volatility by applying a three-month moving average transformation prior to calculating the AUROC statistic. Thus, the ability to capture U.S. business cycle properties that the NBER deems most important appears to be a unique feature of the Bräuning and Koopman (2014) collapsed dynamic factor methodology.

Figure 4 plots the time series of the CFNAI-MA3 and the alternative CFNAI-MA3 with NBER recession shading. Comparing the two indexes in figure 4, we note that the alternative CFNAI-MA3's improvement in AUROC over the CFNAI-MA3 stems largely from its ability to more accurately capture the timing of U.S. recessions prior to 1990. One way in which to see this is to examine periods where the alternative CFNAI-MA3 falls below the dashed line in figure 4. As described in box 5, the ROC framework can also be used to arrive at a single threshold value distinguishing NBER expansions from recessions that equally weights the desire to correctly capture both. The dashed line in the figure is our estimate of this threshold. At –0.7, this threshold for the alternative CFNAI-MA3 is in line with the value first put forth in Evans, Liu, and Pham-Kanter (2002) that has been used as a threshold for the CFNAI-MA3 and slightly above the value computed by Berge and Jordà (2011) using the same ROC methodology (–0.8). Examining values above and below –0.7 during the NBER recessions for the alternative CFNAI-MA3, we note an improvement in AUROC, largely resulting from the fact that it produces fewer false positives and false negatives during the 1969–70, 1973–75, 1980, and 1981–82 recessions. This is the case even though it is slightly ahead of the CFNAI-MA3 in the timing of several remaining recessions.

We can get a sense of what is driving the differences in timing of U.S. recessions across the two measures by breaking down the difference between the CFNAI-MA3 and the alternative CFNAI-MA3 into contributions from the various assumptions of the dynamic factor models building up to the alternative CFNAI-MA3. In essence, this calculation amounts to redisplaying the information

FIGURE 3

Recession and expansion distributions for the alternative CFNAI-MA3



Notes: The three-month moving average of the alternative Chicago Fed National Activity Index (alternative CFNAI-MA3) is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details). The figure displays histograms for the distribution of alternative CFNAI-MA3 values during U.S. recessions (blue) and expansions (red), according to the timing conventions established by the National Bureau of Economic Research, with the bin sizes based on the number of observations in each sample over the period May 1967 through February 2014.
Source: Authors' calculations based on data from Haver Analytics.

TABLE 2

AUROC for NBER recessions and RMSE ratios for current quarter GDP growth predictions

	AUROC	In-sample RMSE ratio	Out-of-sample RMSE ratio
DF	0.95	0.89	0.98
DF-HC	0.95	0.93	0.99
DF-HAC	0.95	0.93	0.85
CDF-HAC	0.98	0.58	0.81

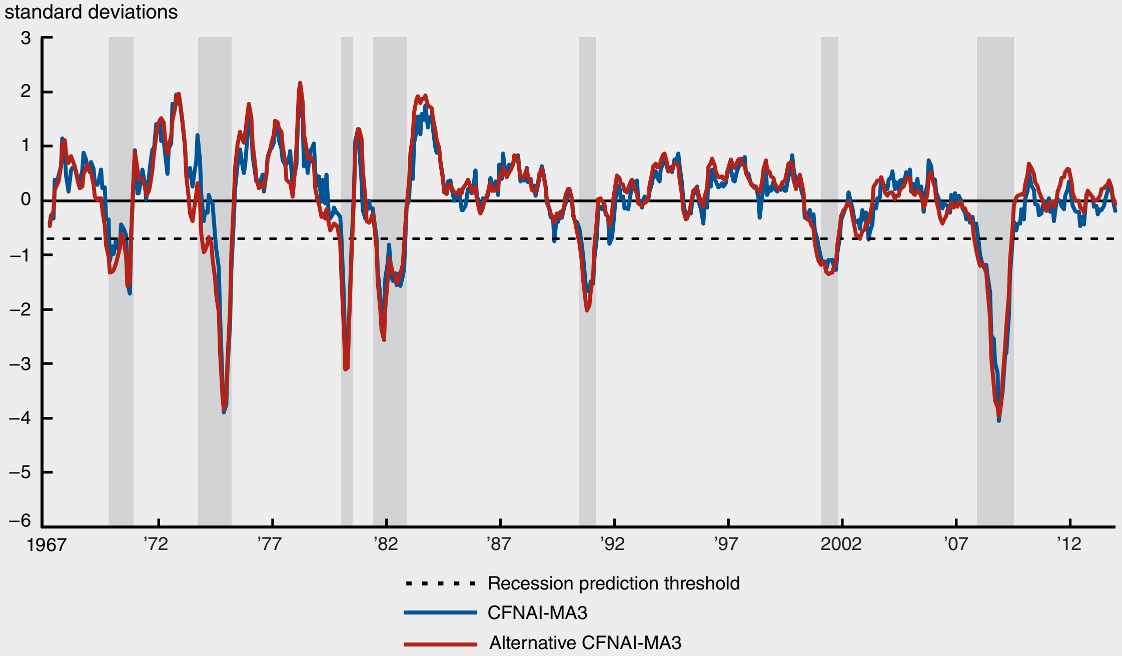
Notes: The table displays areas under the receiver operating characteristic (ROC) curve (AUROC) and root mean squared error (RMSE) ratios for current quarter real gross domestic product (GDP) growth forecasts based on the three-month moving averages of the four dynamic-factor-based alternatives to the Chicago Fed National Activity Index (CFNAI). The four dynamic-factor-based indexes—DF, DF-HC, DF-HAC, and CDF-HAC—are derived from methodologies based on Giannone, Reichlin, and Small (2008), Doz, Giannone, and Reichlin (2012), Jungbacker, Koopman, and van der Wel (2011), and Bräuning and Koopman (2014), respectively (see the text for further details). The closer the AUROC value is to 1, the more accurate a dynamic-factor-based index is in signaling U.S. recessions and expansions as determined by the National Bureau of Economic Research (NBER). An RMSE value of less than 1 indicates a dynamic-factor-based index's forecast that is more accurate than a similar forecast based on the traditional CFNAI using the nowcasting models described in Brave and Butters (2013) for in-sample comparisons over the period March 1967 through February 2014 and Brave and Butters (2010) for out-of-sample comparisons over the period December 2003 through April 2013 (see the text for further details).
Sources: Authors' calculations based on data from the Federal Reserve Bank of Philadelphia, Real-Time Data Set for Macroeconomists; and Haver Analytics.

already presented in figure 2 (p. 27) in a slightly different manner in order to highlight the impact of the cumulative changes across three-month moving averages of the CFNAI and DF-HAC and CDF-HAC indexes (that is, CFNAI-MA3, DF-HAC-MA3, and CDF-HAC-MA3) discussed previously around business cycle turning points. We can decompose the difference between the CFNAI and alternative CFNAI into 1) the difference between the CFNAI and DF-HAC index and 2) the difference between the DF-HAC index and the CDF-HAC index.[11] To arrive at the same measure for the difference between the CFNAI-MA3 and alternative CFNAI-MA3, we take three-month moving averages of all three indexes.

Figure 5 displays our decomposition of the difference between the CFNAI-MA3 and alternative CFNAI-MA3 into two components. The bars in the figure represent contributions to the total difference (represented by the dashed line in the figure) by these two components. The red bars capture the cumulative effect of incorporating dynamics in the static factor model and real

FIGURE 4

CFNAI-MA3 versus alternative CFNAI-MA3



Notes: The figure displays the three-month moving average of the Chicago Fed National Activity Index (CFNAI-MA3) and the alternative CFNAI-MA3, which is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details). Both indexes are shown over the period May 1967 through February 2014. The dashed black line at –0.7 is the recession prediction threshold for the alternative CFNAI-MA3 described in the text. Shading indicates U.S. recessions as identified by the National Bureau of Economic Research.
Source: Authors’ calculations based on data from Haver Analytics.

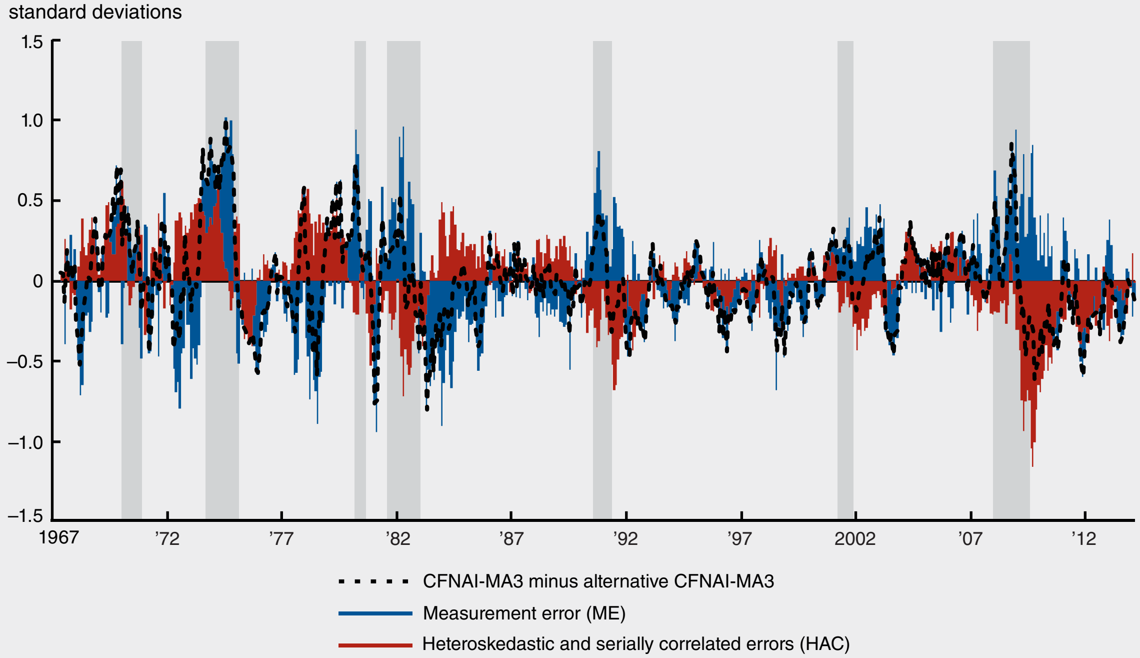
GDP growth along with relaxing the PCA assumptions on the idiosyncratic error structure of the static factor model—that is, the CFNAI-MA3 minus the DF-HAC-MA3. We refer to this component as HAC in the figure as it is primarily the latter feature that dominates the contribution to the total difference. The blue bars capture the marginal effect of the *measurement error* (ME) we estimate in the Bräuning and Koopman (2014) model that arises from bias in the use of PCA—that is, the DF-HAC-MA3 minus the CDF-HAC-MA3. While ME is mean zero by construction over the entire sample period, the large magnitude of many of its realizations in this figure suggests that the CFNAI-MA3 is likely biased. One can see from figures 4 and 5 that HAC primarily accounts for the better fit of the alternative CFNAI-MA3 (relative to the CFNAI-MA3) for the 1969–70 and 1973–75 recessions, while ME is mostly responsible for the improvement in fit for the 1980 and 1981–82 recessions.

More recently, measurement error has begun to play more of a secondary role in explaining the discrepancies between the CFNAI-MA3 and alternative CFNAI-MA3. This has primarily to do with the way in which each index accounts for the protracted weakness of personal consumption and housing indicators during the recovery from the 2007–09 recession and, to a lesser extent, the employment-related indicators as well. The alternative CFNAI-MA3 reinterprets what is due to idiosyncratic drivers of variance in the underlying data series versus what is due to common drivers on the basis of how it has related historically to real GDP growth. For the HAC component to be so strongly negative in figure 5 since 2007 implies that the alternative CFNAI-MA3 indicates growth in economic activity due to personal consumption and housing during the recovery has been greater than what has been indicated by the traditional CFNAI-MA3.

However, since 2007, real GDP growth on average has been weak enough in comparison with the alternative CFNAI-MA3 to suggest that the trend rate of real GDP growth has fallen. This result can be seen in figure 6, with our estimate of the trend rate of real GDP growth decreasing from 2.9 percent in the fourth quarter of 2007 to 2.4 percent in the fourth quarter of 2013. As a point of comparison, the Congressional Budget Office’s (CBO) estimate of potential real GDP growth

FIGURE 5

Accounting for the difference between the CFNAI-MA3 and alternative CFNAI-MA3



Notes: The figure displays the difference between the three-month moving average of the Chicago Fed National Activity Index (CFNAI-MA3) and the alternative CFNAI-MA3, which is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details). Moreover, it shows this difference broken down into contributions from allowing for heteroskedastic and serially correlated errors (HAC) and measurement error (ME) over the period May 1967 through February 2014 (see the text for further details). Shading indicates U.S. recessions as identified by the National Bureau of Economic Research.

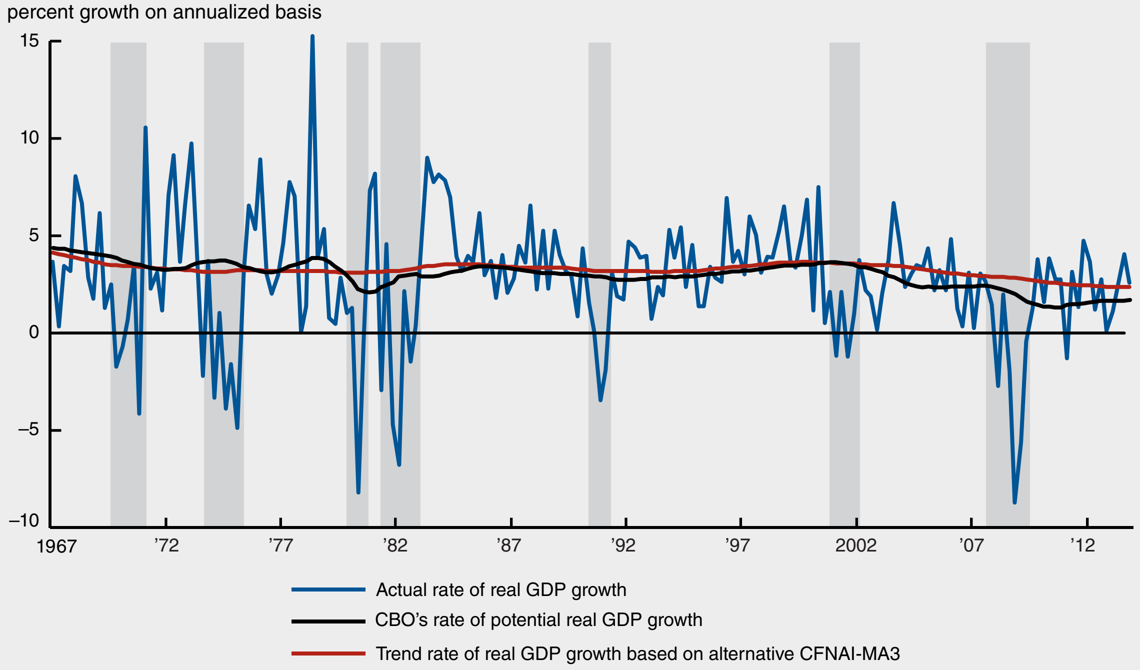
Source: Authors' calculations based on data from Haver Analytics.

is also displayed in figure 6. Our estimate of the decrease in the trend rate of real GDP growth is somewhat smaller than the concurrent change in the CBO's estimate of potential real GDP growth—a decline from 2.4 percent in the fourth quarter of 2007 to 1.7 percent in the fourth quarter of 2013. That said, our estimate of trend GDP growth is also slightly higher than the CBO's estimate of potential real GDP growth for much of the past decade. This is the case even though over the full sample period (1967:Q1–2013:Q4) they exhibit a correlation coefficient of 0.85. The large negative HAC values from personal consumption and housing indicators since 2007 have been often wholly or partially offset by large positive measurement errors from the production and income indicators. This feature of the data has prevented the alternative CFNAI-MA3 from being even further above the CFNAI-MA3 during this period, masking the implied inference for the decline in trend GDP growth.

With recent HAC values near zero or slightly positive (see figure 5), the alternative CFNAI-MA3 suggests that the pervasiveness of the weakness in the household sector is currently more limited than previously thought according to the CFNAI-MA3. This development is a good omen for the continued expansion of the U.S. economy in 2014 if the ongoing recovery in the housing market persists. Recent negative ME values (see figure 5) also suggest that the impact of the weakness in production and income indicators in early 2014 on the CFNAI-MA3 will likely be transitory. While the alternative CFNAI-MA3 also fell into negative territory in February 2014 (see figure 4), it remained much closer to its historical average than the CFNAI-MA3. Using the nowcasting model for real GDP growth described in box 4 (p. 26) as of March 20, 2014, we estimate that real GDP in the first quarter of 2014 increased at an annual rate of 1.8 percent, which is 0.5 percentage points below our current estimate of 2.3 percent for trend real GDP growth. By comparison, the *Blue Chip Economic Indicators* consensus forecast for first quarter real GDP growth on March 10, 2014, was 1.9 percent. In the next section, we evaluate the historical performance of our nowcasting model.

FIGURE 6

Actual versus potential and trend real GDP growth



Notes: The figure displays the annualized quarterly rate of real gross domestic product (GDP) growth, the Congressional Budget Office’s (CBO) estimate of the rate of potential real GDP growth, and our estimate of real GDP’s historical trend rate of growth based on the three-month moving average of the alternative Chicago Fed National Activity Index (alternative CFNAI-MA3), which is constructed using the Bräuning and Koopman (2014) methodology (see the text for further details). All data are plotted over the period 1967:Q1 through 2013:Q4. Shading indicates U.S. recessions as identified by the National Bureau of Economic Research.

Source: Authors’ calculations based on data from Haver Analytics.

Nowcasting real GDP growth

Real GDP is the broadest measure of U.S. economic activity, but it is produced with a significant lag of up to three months. Therefore, linking its current quarter growth rate to the more readily available monthly CFNAI with a nowcasting model has a natural appeal. Furthermore, nowcasts of real GDP growth produced using the CFNAI and similar indexes have been shown to be quite accurate in several instances.[12] To generate nowcasts, we incorporate annualized quarterly real GDP growth into all of our dynamic factor models. However, we show here that only the alternative CFNAI, which is based on the Bräuning and Koopman (2014) methodology, significantly boosts the explanatory power of the dynamic factor model for real GDP growth, further suggesting that the PCA estimate of the index is indeed biased because of a lack of international trade, government spending, and other indicators that inform real GDP growth.

This can be seen in the second column of table 2 (p. 31), which displays in-sample root mean squared error (RMSE) ratios for the nowcasts from the three-month moving averages of the DF, DF-HC, DF-HAC, and CDF-HAC indexes. A number less than 1 indicates an improvement in fit for the quarterly real GDP growth data relative to traditional CFNAI-MA3 nowcasts based on the nowcasting model described in Brave and Butters (2013). While all of the dynamic-factor-based indexes demonstrate an in-sample RMSE ratio of less than 1, the improvement in relative fit for the CDF-HAC index, or alternative CFNAI, at 42 percent dwarfs the others. This is perhaps not surprising given the flexibility of the CDF method in matching the index to observed real GDP growth. A more convincing test of the ability of the Bräuning and Koopman (2014) methodology to correct for potential bias due to a lack of international trade, government spending, and other indicators informing real GDP growth would be to test its ability to nowcast when current quarter real GDP growth is not observed.

As it turns out, the Bräuning and Koopman (2014) methodology is also important for improving the out-of-sample accuracy of our nowcasts, though its relative improvement is not much larger than that achieved with the methodology for the DF-HAC index. To arrive at this conclusion, we estimated three-month moving averages of all four dynamic-factor-based indexes using a real-time archive of the CFNAI data series covering the period December 2003 through April 2013 and the available "vintage" of real GDP growth in those months from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists.[13] We then compared our nowcasts made in the months of each vintage of real GDP growth to the subsequent real-time GDP release being forecasted to compute out-of-sample RMSE ratios similar to the in-sample fits discussed before.[14] As the basis for comparison in this exercise, we used similarly constructed RMSE values based on the within-quarter nowcasting model described in Brave and Butters (2010). The out-of-sample RMSE ratios are shown in the third column of table 2 (p. 31). While all of the ratios are again less than 1, the three-month moving average of the CDF-HAC index (alternative CFNAI) is still the best model in this real-time setting, with a 19 percent improvement in forecast accuracy compared with the Brave and Butters (2010) nowcast.

The differential accuracy of the CDF-HAC-MA3 nowcasts in our real-time out-of-sample nowcasting exercise is not as large in comparison to what we find based on in-sample evidence. This result suggests to us that the advantage provided by allowing for measurement error in the CFNAI-MA3 in nowcasting real GDP growth is somewhat limited given our current nowcasting framework. In fact, when we correlate the forecast errors from our real-time exercise with the real-time contribution of net exports and government spending to real GDP growth, we obtain a correlation coefficient of 0.4. In many ways, however, we are not making full use of the flexibility provided by the Bräuning and Koopman (2014) methodology. In future research, we plan to explore ways in which to improve on our results—by incorporating additional factors, by adding international trade and government spending indicators to the current list of 85, or by employing estimation methods that allow for more informed dynamics and/or parameter shrinkage.

Conclusion

By building on the existing framework of the CFNAI with Bräuning and Koopman's (2014) method of collapsed dynamic factor analysis, we are able to readily extend and improve our existing methodology. Given the resulting alternative CFNAI's superior past performance in predicting current quarter U.S. real GDP growth and very high correlation with NBER recessions, it may very well be a better method to both nowcast real GDP growth and assess the state of U.S. business cycles than the current CFNAI. Bräuning and Koopman's methodology also allows us to address several of the persistent criticisms of the CFNAI, including the problem of overweighting certain sectors of the U.S. economy and the important omissions of certain data series (for example, those concerning international trade and government spending) in nowcasting real GDP growth.

Another benefit of Bräuning and Koopman's (2014) methodology is that it makes it possible to produce both current quarter predictions of real GDP growth and an estimate of the trend rate of real GDP growth with each new index release. As of March 20, 2014, we estimate that real GDP in the first quarter of 2014 increased at an annual rate of 1.8 percent, which is 0.5 percentage points below our current estimate of 2.3 percent for trend real GDP growth. While we are still in the process of investigating the best nowcasting model with which to achieve both of these goals, our work so far suggests that this is a promising direction for future research with the CFNAI. Our analysis here also has implications for the current interpretation of the index. While the alternative CFNAI fell into negative territory in early 2014, it suggests that the pervasiveness of the weakness in the household sector (as well as its drag on U.S. economic activity) is more limited than previously thought according to the traditional CFNAI and that the impact of the recent weakness in the production and income indicators on the index is likely to be transitory.

NOTES

[1]Additional background information on the CFNAI and its method of construction is available at www.chicagofed.org/digital_assets/publications/cfnai/background/cfnai_background.pdf. A complete list of the 85 indicators, their associated categories, and their respective weights in the overall index is available at www.chicagofed.org/digital_assets/publications/cfnai/background/cfnai_indicators_list.pdf.

[2]See the next section and box 1 (p. 21) for details on principal components analysis (PCA) and on how the CFNAI is the first principal component of the 85 data series (that is, the single component common to each data series that explains the most variation across all 85).

[3]The terms *nowcast* and *nowcasting* are derived from combining the words *now* and *forecasting*. Nowcasting techniques are commonly used in economics nowadays because they permit economists today to predict the present (and recent past) of standard measures of the economy (such as real gross domestic product, or GDP), which are often determined after a long delay.

[4]See box 3 (p. 24) for more details on CDF analysis.

[5]See box 1 (p. 21) for further details on the factor model representation of the CFNAI.

[6]Jungbacker, Koopman, and van der Wel (2011) also describe a full maximum likelihood estimator of the collapsed dynamic factor model. In order to make direct comparisons across methodologies, we only consider their EM algorithm estimation method in our article.

[7]Technically, the Stock and Watson (2002) methodology can also incorporate mixed-frequency data. However, because of that methodology's lack of dynamics, this takes place as an additional transformation of the data in its algorithm.

[8]The trend component captures long-run factors, such as potential growth in productivity, capital, and labor. In contrast, the cyclical component captures medium-run factors driving economic growth and is generally associated with the business cycle—the periodic fluctuations in economic activity around its long-term historical trend.

[9]Interestingly, Brave (2008) found similar results for the same two categories (namely, the employment, unemployment, and hours category and personal consumption and housing category) when looking at the impact of slow-moving changes in the average values of their data series over time.

[10]See Brave and Butters (2012a, 2012b) for examples with financial data.

[11]In other words, the difference between the CFNAI and alternative CFNAI is the sum of the difference between the CFNAI and DF-HAC index and the difference between the DF-HAC index and the CDF-HAC index.

[12]See, for instance, Brave and Butters (2010). Other examples using factor models to forecast GDP growth are Stock and Watson (2002) and Giannone, Reichlin, and Small (2008).

[13]This data set is available at www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/.

[14]In making these comparisons, we eliminated quarters where GDP is subject to annual and benchmark revisions.

REFERENCES

Baker, S. G., and B. S. Kramer, 2007, "Peirce, Youden, and receiver operating characteristic curves," *American Statistician*, Vol. 61, No. 4, November, pp. 343–346.

Berge, T. J., and Ò. Jordà, 2011, "Evaluating the classification of economic activity into expansions and recessions," *American Economic Journal: Macroeconomics*, Vol. 3, No. 2, April, pp. 246–277.

Bräuning, F., and S. J. Koopman, 2014, "Forecasting macroeconomic variables using collapsed dynamic factor analysis," *International Journal of Forecasting*, forthcoming.

Brave, S. A., 2008, "Economic trends and the Chicago Fed National Activity Index," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 250, May.

Brave, S. A., and R. A. Butters, 2013, "Estimating the trend rate of economic growth using the CFNAI," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 311, June.

__________, 2012a, "Detecting early signs of financial instability," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 305, December.

__________, 2012b, "Diagnosing the financial system: Financial conditions and financial stress," *International Journal of Central Banking*, Vol. 8, No. 2, June, pp. 191–239.

__________, 2010, "Chicago Fed National Activity Index turns ten—Analyzing its first decade of performance," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 273, April.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson, 1988, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, Vol. 44, No. 3, September, pp. 837–845.

Doz, C., D. Giannone, and L. Reichlin, 2012, "A quasi-maximum likelihood approach for large, approximate dynamic factor models," *Review of Economics and Statistics*, Vol. 94, No. 4, November, pp. 1014–1024.

Durbin, J., and S. J. Koopman, 2012, *Time Series Analysis by State Space Methods*, 2nd ed., Oxford Statistical Science Series, Vol. 38, Oxford, UK: Oxford University Press.

Evans, C. L., C. T. Liu, and G. Pham-Kanter, 2002, "The 2001 recession and the Chicago Fed National Activity Index: Identifying business cycle turning points," *Economic Perspectives*, Federal Reserve Bank of Chicago, Third Quarter, pp. 26–43, available at www.chicagofed.org/digital_assets/publications/economic_perspectives/2002/3qepart2.pdf.

Fisher, J. D. M., 2000, "Forecasting inflation with a lot of data," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, No. 151, March.

Giannone, D., L. Reichlin, and D. Small, 2008, "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, Vol. 55, No. 4, May, pp. 665–676.

Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, UK: Cambridge University Press.

Jungbacker, B., S. J. Koopman, and M. van der Wel, 2011, "Maximum likelihood estimation for dynamic factor models with missing data," *Journal of Economic Dynamics and Control*, Vol. 35, No. 8, August, pp. 1358–1368.

Shumway, R. H., and D. S. Stoffer, 1982, "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, Vol. 3, No. 4, July, pp. 253–264.

Stock, J. H., and M. W. Watson, 2002, "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, Vol. 20, No. 2, April, pp. 147–162.

__________, 1999, "Forecasting inflation," *Journal of Monetary Economics*, Vol. 44, No. 2, October, pp. 293–335.

__________, 1998, "Median unbiased estimation of coefficient variance in a time-varying parameter model," *Journal of the American Statistical Association*, Vol. 93, No. 441, March, pp. 349–358.

Watson, M. W., and R. F. Engle, 1983, "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, Vol. 23, No. 3, December, pp. 385–400.