Vol. 34, No. 2

ECONOMIC REVIEW

1998 Quarter 2

FEDERAL RESERVE BANK
OF CLEVELAND

Vol. 34, No. 2

http://clevelandfed.org/research/review/
Economic Review 1998 Q2

1998 Quarter 2

ECONOMIC REVIEW

1998 Quarter 2
Vol. 34, No. 2

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the *Review* are available through our Corporate Communications & Community Affairs Department. Call 1-800-543-3489 (OH, PA, WV) or 216-579-2001, then immediately key in 1-5-3 on your touch-tone phone to reach the publication request option. If you prefer to fax your order, the number is 216-579-2477.

Economic Review is available electronically through the Cleveland Fed's site on the World Wide Web: *http://www.clev.frb.org*. We also offer a free online subscription service to notify readers of additions to our Web site. To subscribe, please send an email mesage to *econpubs-on@clev.frb.org*.

Editor: Michele Lachman
Design: Michael Galka
Typography: Liz Hanna

ISSN 0013-0281

Bank Diversification: Laws and Fallacies of Large Numbers

by Joseph G. Haubrich

Joseph G. Haubrich is a consultant and economist at the Federal Reserve Bank of Cleveland. He thanks Steve Zeldes, who introduced him to the fallacy of large numbers over a decade ago but who bears no responsibility for the author's subsequent application of that knowledge.

Conventional wisdom states that large banks are safer than small banks because they can diversify more. This conventional wisdom, however, confuses risk with probability of failure. While the *law* of large numbers does imply that a large bank is less likely to fail than a small bank, equating this tendency with lower risk falls into what Samuelson [1963] termed the *fallacy* of large numbers. A $10 billion bank may be less likely to fail than a $10 million bank, but it may also saddle the investor with a $10 billion loss.

In this article, I hope to clarify what this distinction means for banks. Banks diversify by growing—by adding risks—something distinctly different from the subdivision of risk behind standard portfolio theory. A simple mean-variance example will make the point that a risk-averse bank owner need not value diversification by addition. After that, I take a regulator's perspective and consider how a bank guarantee fund, such as the deposit insurance agency, views bank growth and diversification. After a short review of why diversification by adding risks decreases the probability of bank failure, I look at how such diversification alters the expected value of deposit insurance agency payments, then turn to diversification's impact on the deposit insurance agency's expected utility, using recent results from the theory of *standard* risk aversion.

To concentrate on the cleanest example, this article stays with the case of independent and identically distributed risks. This admittedly ignores the alleged ability of large banks to diversify regionally[1] or the possibly adverse incentives of deposit insurance (Boyd and Runkle [1993], Todd and Thomson [1991]).

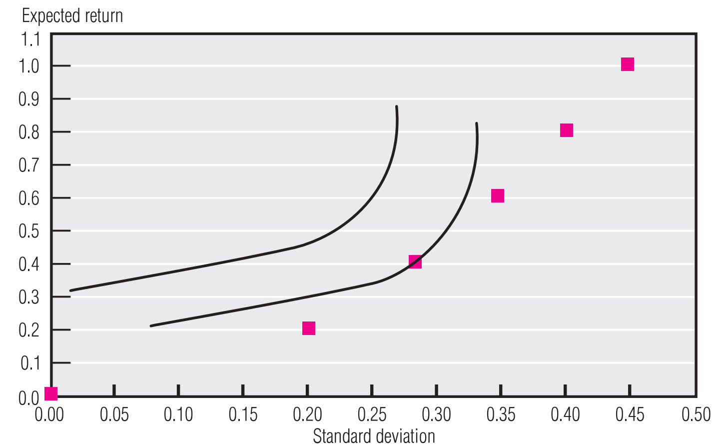
I. A Simple Example

Probably the easiest way to understand the effects of diversification by adding risks is to consider a bank financed exclusively by an owner/investor who cares only about means and variances. With no debt, failure ceases to be an issue; instead, the question is the utility-maximizing portfolio for the bank's owner.

1 Compare Haubrich (1990) with Kryzanowski and Roberts (1993). Even small banks may diversify, however, by selling loans or participating in mortgage pools or other forms of securitization.

FIGURE 1

Bank Opportunity Set



SOURCE: Author.

The owner and sole equity holder has, conveniently for us, sunk his entire wealth W into the bank. He faces the problem of dividing his portfolio between holding S safe government bonds with a certain return of zero and investing in some number K of risky, independent bank loans with returns R_i normally distributed as $N(\mu,\sigma^2)$, that is, with mean μ and variance σ^2. If each loan costs a dollar, the investor's budget constraint is $W = S + K$. These bank loans are indivisible—the bank cannot diversify by spreading one dollar across many loans. Then the return on the portfolio is

$$\tilde{R}_p = \frac{\sum_{i=1}^{K} R_i}{W}.$$

Since $\sum_{i=1}^{K} R_i$ is distributed $N(K\mu, K\sigma^2)$, the portfolio has expected return $E(\tilde{R}_p) = \frac{K}{W}\mu$ and variance $\sigma^2(\tilde{R}_p) = \frac{K}{W^2}\sigma^2$. From this, simple substitution (for this and other standard techniques, see Fama and Miller [1972], chapter 6, section IV) implies that

$$(1) \qquad E(\tilde{R}_p) = \frac{\mu}{\sigma}\sqrt{K}(\tilde{R}_p).$$

In mean-standard deviation space, equation (1) defines a portfolio opportunity set, or the different risk and return combinations available to the investor. This set is illustrated by figure 1 (for $W = 5$, $\mu = 1$, and $\sigma = 1$). The opportunity set is disjointed, since the decision to add another loan is discrete. Depending on the shape of the indifference curves, the bank owner may choose to put none, all, or some of his wealth into bank loans. Figure 1 shows a typical case with an interior solution, illustrating quite clearly that the bank does not always wish to diversify. Stated another way, the portfolio return is distributed $N(\frac{K}{W}\mu, \frac{K}{W^2}\sigma^2)$, so that as the bank invests in more loans, the standard deviation increases as well as the expected return. Preferences determine which of them matters more.

An all-equity bank offers a nice illustration, but does not provide a very representative case. Even a stylized bank should have deposits.

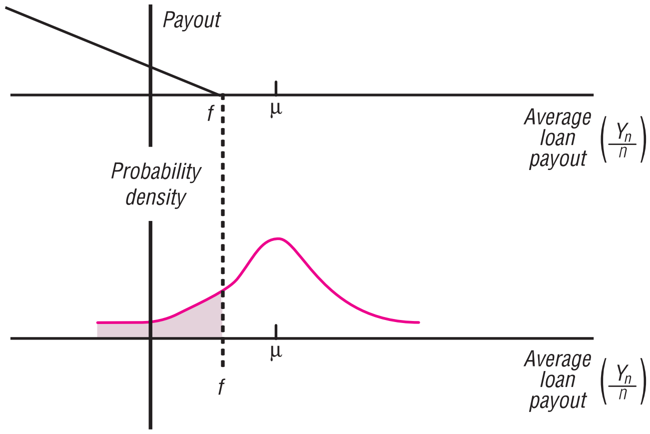
II. Should the Deposit Insurance Agency Want Banks to Diversify?

Allowing banks to take in deposits means allowing banks to fail. The return on assets may not cover the payments promised to depositors. Many countries provide some sort of deposit insurance, which guarantees the deposit. In the case where bank assets cannot cover the payments promised to depositors, the difference becomes a liability for the deposit insurance agency (which may be either public or private). This provides a natural focal point for our discussion, although what happens in reality is much more complicated. Actual banks raise money in many different ways, using several types of preferred stock, subordinated bonds, commercial paper, and both insured and uninsured deposits. What happens in bankruptcy is at best complicated and at worst unknown, because the courts must determine the validity of claims as diverse as offsetting deposits and the source-of-strength doctrine. A detailed consideration of how each class of investors views diversification, then, is beyond the scope of this article. Instead, to make what is admittedly a simple point, I concentrate on the deposit insurance agency, which ultimately bears the liability for bank failures.

The deposit insurance agency steps in if the realization of bank assets Y is too low to repay the face value of the debt F, that is, if $Y < F$. This is a fairly general formulation in that the assets producing Y may be funded by means other than deposits, but it is not completely general because it ignores the possibility that the deposit insurance agency may have priority over some investors. For the rest of the article, however, I will restrict myself to purely deposit-financed banks. This is at the opposite extreme from the discussion in section I, and hence provides a nice comparison.

FIGURE 2

Payout Function and Probability



SOURCE: Author.

What is the face value of the debt, F? With no capital, if the bank funds n projects, each requiring funds f, the face value is the sum of the deposits, $F = n \cdot f$. The payout of bank assets is likewise the sum over the different projects,

$$Y_n = \sum_{i=1}^{n} x_i ,$$

where n indexes the number of projects in which the bank has invested.

The Probability of Bank Failure

How likely is it that this bank will fail? The answer is $Pr\,(\sum_i x_i < F)$ or

(2) $\quad Pr\,(Y_n < n \cdot f\,).$

Assume that the x_i's are independent and identically distributed (i.i.d.), with mean $E\,(x_i) = \mu$; further assume that $f < \mu$, so that the face value of the debt is smaller than the expected payout of the assets.

We can rewrite expression (2) as

(3) $\quad Pr\,(\frac{Y_n}{n} < f\,)$

because the set $\{y{:}y < n \cdot f\,\}$ is the same as the set $\{y{:}\,\frac{y}{n} < f\,\}$.

The weak law of large numbers (see Shirayev [1984], theorem 2, p. 323; for a more elementary discussion, see Hogg and Craig [1978, chapter 5]) tells us that provided $E\,|x_i| < \infty$ and $E\,x_i = \mu$, then for all $\varepsilon > 0$,

$$Pr\,\{|\frac{Y_n}{n} - \mu| \geq \varepsilon\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In particular, since $f < \mu$, $Pr\,(\frac{Y_n}{n} < f\,)$ $< Pr\,(|\frac{Y_n}{n} - \mu| \geq \mu - f\,)$. That is, we can represent the values $\frac{Y_n}{n}$ below f as values that are more than $\mu - f$ away from the mean μ. Thus, as Diamond (1984) explicitly states, the weak law of large numbers implies that diversification by adding risks reduces the probability of bank failure.[2]

The Expected Value of the Deposit Insurance Agency's Liabilities

As Samuelson points out, a rational utility maximizer maximizes expected utility, not the probability of success. The probability of each outcome must be weighted by the utility of that outcome. As mentioned before, the failure of a \$10 billion bank may cost the deposit insurance agency more to resolve than that of a \$10 million bank.

In the simplest case of risk neutrality, expected utility corresponds to expected value. The first question, then, equivalent to assuming risk neutrality on the agency's part, concerns the expected value of the deposit insurance agency's payout.[3] Determining the expected payout value becomes a question of finding the expected value of a particular function. The deposit insurance agency must pay

$$(4) \quad \begin{cases} 0 & \text{if } Y_n \geq F, \text{ that is, if } \frac{Y_n}{n} \geq f \\ \text{and} & \\ F - Y_n & \text{if } Y_n < f, \text{ that is, if } \frac{Y_n}{n} < f. \end{cases}$$

Figure 2 plots the function along with a typical density function.

It is worth noting that the expected value of (4) is not a conditional expectation. If the set $A = \{Y_n{:}\,Y_n < F\,\}$, then the expected value of (4) is $P\,(A)\,E\,(Y_n\,|\,A)$ rather than $E\,(Y_n\,|\,A)$. A simple example will make this clear. Take a four-point distribution, $P\,(1) = P\,(2) = P\,(3) = P\,(4) = \frac{1}{4}$. Then $E\,(X) = \frac{1}{4}(1 + 2 + 3 + 4) = \frac{5}{2}$. Now define the function $g(x)$ as $g(x) =$ $\{0$, if $x \geq 2.5$, and x, if $x < 2.5\}$. Then $E\,[g(x)]$ $= \frac{1}{4}(1 + 2) = \frac{3}{4}$, while $E\,[x\,|\,x \leq 2.5] = \frac{3}{2}$.

2 See also Winton (1997).

3 Although the calculation is not particularly difficult, I have not seen it before in the literature.

The question before us is what happens to the expected value of the deposit insurance agency's payments as the bank diversifies. Recall that the deposit insurance agency pays off if $\frac{Y_n}{n} < f$ or, equivalently, $\frac{\sum_{i=1}^{n} x_i}{n} < f$. By the strong law of large numbers, the mean of the partial sums $\sum_{i=1}^{n} x_i$ converges to a mass point on $E(x)$; that is, the sample means approach the true mean. Intuition suggests that the expected value of anything below the mean (and, *a fortiori*, anything below f) will have very little importance, that is, an expected value approaching zero. Put another way, as the bank gets very large the probability gets vanishingly small, and the average loan payoff falls below the amount promised to depositors; thus, the probability of a deposit insurer having to make a payoff gets so low that the expected value of that payoff approaches zero.

To establish this rigorously and to understand what diversification does to the expected value of the deposit insurance agency's payments requires a more formal approach, which is provided in the appendix. The intuition and results are less complicated, however. As a bank makes more loans, the expected value of deposit insurance agency payouts tends toward zero, and so the deposit insurance agency would like to encourage large banks. Diversification by adding loans works.

III. A Risk-Averse Deposit Insurance Agency

When risk aversion enters the picture, however, a deposit insurer can be worse off with larger banks. Strictly speaking, what Samuelson terms the fallacy of large numbers enters only with risk aversion. Applying it to an organization such as a deposit insurance agency, rather than to an individual, requires some justification. One possibility is that a publicly sponsored deposit insurance agency must obtain its funds by taxing people, either indirectly through its assessment on banks or directly by Congressional appropriation. Risk aversion by the deposit insurance agency may then reflect risk aversion on the part of those taxed, or nonlinearities associated with distortionary taxation. Alternatively, the risk aversion may result from the incentives, constraints, and information facing the organization: The managers running it may act risk averse, perhaps because their future income depends on their performance. (Of course, as Kane [1989] points out, this dependency may sometimes promote risk-seeking behavior, as in the FSLIC case.)

Conditions for the Fallacy

Samuelson (1963) shows that if a consumer rejects a bet at *every* wealth level, then he will always reject any independent sequence of those bets. Under the Samuelson condition, if the deposit insurance agency found one bank loan too risky, it would find a portfolio of any number too risky. It would be no happier to insure a large bank with many loans than a small bank with few loans.

Samuelson posits a rather stringent condition. It rules out, for example, constant relative risk-averse (CRRA) utility, because CRRA exhibits decreasing absolute risk aversion (DARA), and so some unacceptable gambles would become acceptable at higher wealth levels. Pratt and Zeckhauser (1987, p. 143) improve considerably on the condition with their notion of *proper* risk aversion. The conditions for proper risk aversion answer the question, "An individual finds each of two independent monetary lotteries undesirable. If he is required to take one, should he not continue to find the other undesirable?" In our problem, if the deposit insurer does not like the risk in a bank with one loan, then it will not like the risk in a bank with two loans. Proper risk aversion shares one defect with Samuelson's condition, however: It is difficult to characterize and difficult to determine whether a particular utility function satisfies the condition.

A slightly stronger condition with a simple characterization is proposed by Kimball (1993), whose *standard* risk aversion implies proper risk aversion. It thus applies a slightly stronger condition than is strictly necessary for the fallacy. If a utility function displays standard risk aversion, then an investor who dislikes a bet will also dislike a collection of such bets.

Kimball (1993) shows that necessary and sufficient conditions for standard risk aversion are (monotone) DARA and (monotone) decreasing absolute prudence. If the utility function in question has a fourth derivative, then these conditions (where, as before, W indicates a person's wealth) become

$$(5) \quad \frac{d}{dW}\left(-\frac{u'}{u''}\right) < 0 \text{ or } u^{(3)} > \frac{(u'')^2}{u'} > 0$$

and

$$(6) \quad \frac{d}{dW}\left(-\frac{u^{(2)}}{u^{(3)}}\right) < 0 \text{ or } u^{(4)} < \frac{(u^{(3)})^2}{u^{(2)}} < 0.$$

A key point here is that the individual finds each independent risk undesirable. (Kimball has a slightly weaker, more technical condition that he calls loss aggravation.) This certainly applies to the problem as we have defined it, because the payoff to the deposit insurance agency is nonpositive—at best, it pays nothing. This is not the only way to structure the problem, however, because the deposit insurance agency collects premiums from banks. A major strand in banking research has been to ascertain whether the insurance premiums are fairly priced, that is, whether they represent a tax or a subsidy on the bank (Pennacchi [1987], Thomson [1987]). The empirical results are mixed, varying by time period and by bank; in any case, they assume risk neutrality and so do not directly answer the question most relevant here. It makes sense, then, to think about both possibilities—the case where the deposit insurance agency finds insuring a single loan undesirable and the case where it finds insuring a single loan desirable.

In the first case, where the deposit insurance agency dislikes insuring an individual loan, expressions (5) and (6) provide sufficient conditions for the agency to dislike insuring any portfolio of such loans. That is, diversification by adding risks does not work; adding risks makes the insurance agency (guarantee corporation) worse off.

In the second case, where the deposit insurance agency likes insuring an individual loan, equations (5) and (6) do not help. Their derivation presupposes that the agency dislikes the risk it bears. For *favorable* bets, Diamond (1984) builds on Kihlstrom, Romer, and Williams (1981) to develop sufficient conditions for when the fallacy of large numbers is not a fallacy.

Diamond poses the problem in terms of risk premiums and notes that adding risks provides true diversification if it reduces the risk premium. That is, diversification works if the incremental premium for adding the second risky loan to the portfolio is lower than for adding the first (identical) risky loan. Kihlstrom, Romer, and Williams show how to handle risk aversion with two sources of uncertainty by defining a new utility function, given initial wealth W_0 and initial risky bet $\tilde{x}_1$, as

$$(7) \qquad v(x_2) = Eu(W_0 + \tilde{x}_1 + x_2).$$

Now $v(x_2)$, as defined in equation (7), can be treated as a utility function, so Diamond's question comes down to whether $u(\cdot)$ is more risk averse than $v(\cdot)$. If it is, then the risk premium for bearing the second risk is lower than for the first, and the fallacy of large numbers is not a fallacy.

Diamond derives two sufficient conditions under which $u(\cdot)$ will be more risk averse than $v(\cdot)$ Using Jensen's inequality, he shows that

$$(8) \qquad u^{(3)} > 0$$

and

$$(9) \qquad u^{(4)} > 0$$

are sufficient conditions when the risk has zero expected value. When the risk is freely chosen, he must append decreasing absolute risk aversion, equation (5), to these conditions. The reason is that a freely chosen gamble increases mean wealth, which requires us to augment the sufficient conditions.

Notice that inequalities (6) and (9) cannot hold simultaneously because (6) demands a negative fourth derivative and (9) demands a positive fourth derivative. The inequalities apply in different situations, however. Inequality (6) concerns unfavorable bets and describes when bearing one such risk makes the agent less willing to bear another. Inequality (9) concerns favorable bets and describes when bearing such a risk makes the agent even more willing to bear another. The conditions really answer two quite different questions. Since each inequality provides a sufficient but not necessary condition, any contradiction between the answers is more apparent than real.

An important caveat is that this is consciously a partial equilibrium analysis, concentrating on the risk of a single bank. If that bank grows by absorbing smaller ones, the total number of loans insured by the system does not change. A bank merger does not change the total loans insured by the agency, but merely redistributes them. In a bank with many loans, the profitable loans may offset the unprofitable, lessening the guarantor's liability. Since the deposit insurance agency does not share in the positive profits, it cannot undertake a similar offset if the loans are in different banks. In an extreme case, if each bank had only one loan, the insurer would make payments on every nonperforming loan. If all loans were in one bank, the insurer would make payments only if the aggregate loan loss were too large.

This is not the only scenario, however. The bank may grow at the expense of nonbank intermediaries or by making loans that would not be made without the guarantee. Either case results in an increase of total loans guaranteed by the deposit insurance agency, increasing its liabilities as it takes on new loans that must get insured.

TABLE 1

Premium Computation

Risk aversion, a	Face value	π_1	π_2
0.1	1	0.006	0.018
1.0	1	0.066	0.193
10.0	1	0.392	0.979

SOURCE: Author.

An Exponential Example

A simple example can serve to illustrate some of the subtleties involved. To show what can happen, I use an exponential utility function and an exponential distribution. The exponential distribution keeps the algebra simple because sums of exponentials are gamma distributions.[4] Exponential utility exhibits constant, rather than decreasing, absolute risk aversion. It does not satisfy the sufficiency conditions of Kimball ([5] and [6]) or of Diamond ([8] and [9]).

Whether diversification helps or hurts depends on the risk premium. If the risk premium decreases as the investor adds i.i.d. risks, diversification helps. If the risk premium increases, diversification hurts. The simplicity of the example allows us to calculate the risk premium explicitly.

Recall from equation (4) that, for one loan, the deposit insurance agency pays nothing if the loan's payoff exceeds its face value; otherwise it pays the difference. Denoting this function by $g(x)$ (as in the appendix), the risk premium is defined as the π_1 that satisfies

$$(10) \quad u\{W_0 - E[g(\tilde{x})] - \pi_1\} = Eu[W_0 - g(\tilde{x})].$$

With x following the simplest exponential distribution, e^{-x}, the expected value in (10) becomes

$$Eg(\tilde{x}) = (f - 1) + e^{-f}.$$

Using exponential utility of the form e^{-aW} allows us to solve for π_1:

$$(11) \quad \pi_1 = (f - 1 + e^{-f}) - \frac{1}{a}\log[e^{-f} - \frac{1}{1+a}(e^{-f} - e^{af})].$$

For two loans, $g(x)$ is zero if x exceeds $2f$ and $2f - x$ otherwise. The random variable x, as the sum of two independent exponentials, has a gamma distribution,

$$x \sim \frac{xe^{-x}}{\Gamma(2)} = \frac{x}{2}e^{-x}.$$

The expected value then becomes $Eg(\tilde{x}) = 2(f - 1) + 2(f + 1)e^{-2f}$. Solving for the risk premium implicitly defined by $u\{W_0 - E[g(\tilde{x})] - \pi_2\} = Eu[W_0 - g(\tilde{x})]$ using $\pi_2 = Eg(x) - \frac{1}{a}\log Ee^{ag(x)}$ yields

$$(12) \quad \pi_2 = 2(f - 1) + 2(f + 1)e^{-2f} - \frac{1}{a}\log \{e^{-2f}\frac{1}{1+a^2}e^{2af} [1 - (1 + 2f(1 + a))e^{-(1+a)f}] + (1 + 2f)e^{-2f}\}.$$

To complete the example, set f, the face value of the debt, to 1, and compute the premium for several values of risk aversion, evaluating (11) and (12).

The example in table 1 illustrates that diversification does not work in every case. The required risk premium for two loans is higher than for only one: It even exceeds twice the risk premium for one loan. As risk aversion increases, the risk premiums also increase. Although conditions (5) and (6) are not satisfied, the deposit insurance agency dislikes adding more independent risks to its portfolio.

IV. Conclusion

Discussions of banking have been obscured by a false analogy with portfolio theory. A bank diversifies differently than does a mutual fund, adding risks rather than subdividing them. Using the weak law of large numbers to establish that diversified banks have a lower expected failure rate neglects the deeper question of whether this represents a decrease in economic risk. To clearly pose that question is the main point of this article.

Just because a bank is less likely to fail, it is not necessarily less risky. If the insurer, or owner, is risk neutral, a more complicated argument shows that the bank is less risky in the sense of expected value. With risk aversion, however, the question becomes ambiguous. As a practical matter, sufficient conditions exist, and the combination of exponential utility with exponential distributions provides a tractable framework for further exploration.

4 In general, bank loans are more likely to be negatively skewed, while the exponential is positively skewed, so this example is not meant as a realistic description of actual returns.

Appendix

Let each random variable be defined on the probability space (Ω, F, P) and identify Ω with R, the real numbers, without loss of generality. The random variables are then functions on this space, $X_i(\varpi)$, and define $Z_n(\varpi)$ as

$$Z_n(\varpi) = \sum_{i=1}^{n} \frac{X_i(\varpi)}{n}.$$

Next, define the function $g(\varpi)$ as

$$g(\varpi) = \begin{cases} f - X(\varpi) & \text{if} \quad X(\varpi) \leq f \\ \text{and} & \\ 0 & \text{if} \quad X(\varpi) > f. \end{cases}$$

Note that we can think of the expectation $E[X(\varpi)]$ as a random variable, and so $g(E[X(\varpi)] = g(\mu) = 0$, since f < μ. Further define $g_n(\varpi)$ as $g[Z_n(\varpi)]$.

The value of diversification can then be expressed by saying that as n approaches infinity, the expected value of $g(Z_n)$ approaches zero, or

$$\text{(A1)} \quad \lim_{n \to \infty} \int g_n(\varpi) = g(\mu) = 0.$$

To prove (A1), we use Lebesgue's dominated convergence theorem (Royden [1968], p. 229), which says that if $h(\varpi) \geq 0$ is integrable, if $|g_n(\varpi)| \leq h(\varpi)$, and if $g_n(\varpi) \overset{a.s.}{\to} g(\varpi)$, then

$$\lim_{n \to \infty} \int g_n(\varpi) = g(\varpi).$$

The theorem first requires that we prove $g_n(\varpi) \overset{a.s.}{\to} g(\mu)$. To do so, we use the strong law of large numbers for i.i.d random variables (see Breiman [1992, p. 52, theorem 3.30]), which says that for i.i.d. X_1, X_2, X_3 ..., if $E|X_1| < \infty$ then $\frac{\Sigma X_i}{n} \overset{a.s.}{\to} E(X_1)$, where $\overset{a.s.}{\to}$ denotes almost sure convergence, that is, convergence on all but a set of measure (probability) zero.

Hence, given an ϖ, except for a set of measure zero, we have that for any $\varepsilon > 0$, there exists an N such that if $n > N, |Z_n(\varpi) - \mu| < \varepsilon$. Choose $\varepsilon < \mu - f$, which implies that if $|Z_n(\varpi) - \mu| < \varepsilon$, then $Z_n(\varpi) > \mu - \varepsilon > f$. This, with the definition of g, in turn implies that $g_n(\varpi) = 0$. For this ϖ, then, $g_n(\varpi) = g(\mu) = 0$, and, *a fortiori*, $|g_n(\varpi) - g(\mu)| < \varepsilon$. Since $g_n(\varpi) \to g(\mu)$ for each ϖ where $Z_n \to \mu$, the almost sure convergence of the strong law implies the almost sure convergence $g_n(\varpi) \overset{a.s.}{\to} g(\mu)$.

All that remains to be shown is the existence of the integrable bound $h(\varpi)$. For this, use $|X_i(\varpi) + \mu - f|$, which bounds g_n and is integrable because $E|X_1| < \infty$ is a hypothesis of the strong law. Hence, Lebesgue's dominated convergence theorem applies.

As a bank makes more loans, the expected value of deposit insurance agency payouts tends toward zero. Diversification works.

References

Boyd, John H., and David E. Runkle. "Size and Performance of Banking Firms: Testing the Predictions of Theory," *Journal of Monetary Economics*, vol. 31, no. 1 (February 1993), pp. 47–67.

Breiman, Leo. *Probability*, Classics in Applied Mathematics, vol. 7. Philadelphia: Society for Industrial and Applied Mathematics, 1992.

Diamond, Douglas W. "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, vol. 51, no. 3 (July 1984), pp. 393–414.

Fama, Eugene F., and Merton H. Miller. *The Theory of Finance*. Hinsdale, Ill.: Dryden Press, 1972.

Haubrich, Joseph G. "Nonmonetary Effects of Financial Crises: Lessons from the Great Depression in Canada," *Journal of Monetary Economics*, vol. 25, no. 2 (March 1990), pp. 223–52.

Hogg, Robert V., and Allen T. Craig. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.

Kane, Edward J. *The S&L Insurance Mess: How Did It Happen?* Washington, D.C.: The Urban Institute Press, 1989.

Kihlstrom, Richard E., David Romer, and Steve Williams. "Risk Aversion with Random Initial Wealth," *Econometrica*, vol. 49, no. 4 (July 1981), pp. 911–20.

Kimball, Miles S. "Standard Risk Aversion," *Econometrica*, vol. 61, no. 3 (May 1993), pp. 589–611.

Kryzanowski, Lawrence, and Gordon S. Roberts. "Canadian Banking Solvency, 1922–1940," *Journal of Money, Credit, and Banking,* vol. 25, no. 3, pt. 1 (August 1993), pp. 361–76.

Pennacchi, George G. "A Reexamination of the Over- (or Under-) pricing of Deposit Insurance," *Journal of Money, Credit, and Banking,* vol. 19, no. 3 (August 1987), pp. 340–60.

Pratt, John W., and Richard J. Zeckhauser. "Proper Risk Aversion," *Econometrica,* vol. 55, no. 1 (January 1987), pp. 143–54.

Royden, H. L. *Real Analysis,* 2nd ed. New York: Macmillan, 1968.

Samuelson, Paul A. "Risk and Uncertainty: A Fallacy of Large Numbers," *Scientia,* vol. 57 (April–May 1963), pp. 1–6.

Shirayev, A. N. *Probability,* trans. Ralph P. Boas, Graduate Texts in Mathematics, vol. 95. New York: Springer–Verlag, 1984.

Thomson, James B. "The Use of Market Information in Pricing Deposit Insurance," *Journal of Money, Credit, and Banking,* vol. 19, no. 4 (November 1987), pp. 528–37.

Todd, Walker F. , and James B. Thomson. "An Insider's View of the Political Economy of the Too Big to Fail Doctrine," *Public Budgeting and Financial Management: An International Journal,* vol. 3, no. 3 (1991), pp. 547–617.

Winton, Andrew. "Competition among Financial Intermediaries When Diversification Matters," *Journal of Financial Intermediation,* vol. 6, no. 4, (October 1997), pp. 307–46.

Regional Variations in White–Black Earnings

by Charles T. Carlstrom and Christy D. Rollow

Charles T. Carlstrom is an economist at the Federal Reserve Bank of Cleveland and Christy D. Rollow is a research associate at the Federal Reserve Bank of Richmond.
The authors thank Ben Craig and Mark Schweitzer for many helpful discussions.

Between 1960 and 1980, the United States made significant progress in narrowing the earnings gap between whites and blacks. Unfortunately, the last two decades have not been as successful. In 1997, the median income for whites was $47,100; for blacks, it was a much lower $26,500. This difference is not uniform across the country, however. Studies show that although blacks have made the largest strides in the South, the pay gap is still widest there (Murphy and Welch [1990]). This article examines some of the major reasons why blacks' earnings continue to lag whites' and why this problem is especially acute in the southern states.

In the broadest sense, we hope that a greater understanding of the factors driving white–black pay differentials at the regional level will help explain some of the disparities at the national level. Such knowledge is clearly applicable to a wide variety of important public policy issues. In the narrower sense of monetary policy, we seek to enlarge the knowledge base on which policy decisions are made. The ability to explain patterns of labor compensation is central to many debates about labor markets, productivity growth, and inflation dynamics. Furthermore, the regional nature of much discussion surrounding real-time policy decisions—a notable public example being the Federal Reserve System's regularly published *Beige Book*—suggests the importance of understanding how earnings patterns across groups and geographic areas compare with those typically found in aggregate analyses.

Previous studies of earnings disparities between groups have followed Oaxaca (1973), which decomposes wage variations into their explained and unexplained components. Under the null hypothesis, explained differences are typically interpreted as reflecting differences in productivity, while unexplained differences are loosely associated with discrimination. These studies account for regional variations through the use of dummy variables—a method that presumes identical wage-determination processes across all regions. As a result, standard wage decompositions provide aggregated measures, but only limited information about earnings and discrimination within particular areas of the country.

This article differs from earlier studies in that it focuses specifically on decomposing regional wage differentials. The regions considered here are the South and non-South. Our approach is a natural extension of the Oaxaca decomposition. Running separate wage regressions for whites and blacks is equivalent to interacting race with the independent variables. Similarly, estimating separate regressions for the South and non-South enables the researcher to analyze white–black wage differentials without assuming the same wage equation for both regions.

The Oaxaca decomposition has several limitations. While unexplained differences are generally thought to measure discrimination, they may not do so accurately. Likewise, explained differences may be due to discrimination rather than to differences in productivity. Therefore, caution must be used in classifying the reasons behind existing wage disparities. A good approach is to look at our findings as a piece of a much larger puzzle that policymakers and economists can use to better understand the determinants of earnings.

The article proceeds as follows: Section I introduces the Oaxaca decomposition used for identifying white–black wage differentials. Section II discusses the limitations of this approach. To help understand the empirical results covered later, section III provides an overview of discrimination theory. The final sections then specify and estimate U.S. and regional wage models, discuss these results in light of discrimination theory, and summarize our key findings regarding the determinants of wage differentials and the nature of the earnings gap.

I. Decomposition Methodology

One of the reasons for developing the Oaxaca decomposition was to quantify the amount of discrimination in the marketplace. Market discrimination occurs when individuals with the same productivity are paid different wages. Productivity, however, is not observable.

For discrimination to be appropriately captured by the Oaxaca decomposition, the following conditions must be satisfied: 1) The factors that determine productivity (education, for example) can be completely identified; and 2) these factors are accurately measured. Given these conditions, one can estimate what portion of the wage differential is due to productivity and what portion is due to discrimination.

The Oaxaca approach involves estimating separate wage equations for different groups, such as whites and blacks, and then decomposing the pay gaps into explained and unexplained portions. Explained differences are said to be attributable to productivity factors, such as experience and education, while unexplained differences are said to result from discrimination, as already defined. The conceptual framework for this analysis is presented below.

The average wage gap between whites and blacks is given by

$$(1) \qquad Gap \approx \ln(\hat{W}_W) - \ln(\hat{W}_B),$$

where the carets over the W's represent the average wage of whites and blacks. Equation (1), then, is simply the percentage difference between white and black wages. Regression equations for whites and blacks are defined as

$$(2) \qquad \begin{aligned} \ln(W_W) &= X_W \beta_W + \varepsilon \\ \ln(W_B) &= X_B \beta_B + \varepsilon, \end{aligned}$$

where W represents wages, X is the vector of the regressors, and β is the corresponding vector of estimated coefficients (Oaxaca and Ransom [1994]). Within this framework, the geometric means (approximately equal to the arithmetic means) of the equations are as follows (β's represent the ordinary-least-squares [OLS] estimates from equation [2]):

$$(3) \qquad \begin{aligned} \ln(\hat{W}_W) &\approx \hat{X}_W \beta_W \\ \ln(\hat{W}_B) &\approx \hat{X}_B \beta_B . \end{aligned}$$

To understand the earnings disparity between whites and blacks better, it is important to distinguish between differences that can be explained by variations in the explanatory variables and differences that are unexplainable.

There are two ways of decomposing equation (3) into its explained and unexplained components. The first decomposition, referred to as "whites as base," assumes that in the absence of discrimination, the wage structure (and hence productivity) for both races is given by the estimated coefficients for whites.

$$(4) \qquad \begin{aligned} Gap_W &\approx \ln(\hat{W}_W) - \ln(\hat{W}_B) \approx \hat{X}_W \beta_W - \hat{X}_B \beta_B \\ &= \hat{X}_W \beta_W - \hat{X}_B \beta_B + \hat{X}_B \beta_W - \hat{X}_B \beta_W \\ &= (\hat{X}_W - \hat{X}_B)\beta_W + \hat{X}_B(\beta_W - \beta_B). \end{aligned}$$

The alternative ("blacks as base") assumes that the true relationship governing productivity for both groups is given by the regression estimates for blacks.

$$(5) \qquad \begin{aligned} Gap_B &\approx \ln(\hat{W}_W) - \ln(\hat{W}_B) \approx \hat{X}_W \beta_W - \hat{X}_B \beta_B \\ &= \hat{X}_W \beta_W - \hat{X}_B \beta_B + \hat{X}_W \beta_B - \hat{X}_W \beta_W \\ &= (\hat{X}_W - \hat{X}_B)\beta_B + \hat{X}_W(\beta_W - \beta_B). \end{aligned}$$

By construction, the total wage gap between whites and blacks is independent of the method selected. The portions that fall under "explained" and "unexplained" may be different, however. The first term of equations (4) and (5) is the explained component. It is represented by differences in the means of the explanatory variables, *X*, and measures the part of the wage difference that arises because blacks are underrepresented in high-paying, skilled occupations or have less education on average than whites. These differences are said to be "explained" because they represent the wage differential that stems from differences in productivity arising from observed differences in education, tenure, experience, occupation, and so on.

The second term of equations (4) and (5) is the unexplained, or potentially discriminatory, portion. It arises whenever the returns to whites and blacks differ for given values of an explanatory variable, including the intercept. Because measurable characteristics and job qualities account for only a part of the wage gap, the remaining differential is often assumed to result from discrimination.

II. Does the Oaxaca Decomposition Capture Discrimination?

The Oaxaca decomposition employs many assumptions designed to measure discrimination. Unfortunately, in the real world, some of these may be invalid. It is important to understand why the explained and unexplained components of the Oaxaca decomposition may not capture productivity differences and discrimination.

One reason for including explanatory variables such as female, firm size, and region is that they empirically affect wages. Another reason is that differences in the means of some explanatory variables may reflect discrimination rather than individual choice. For example, blacks may be denied promotions or be "segregated" into lower-paying occupations and industries.[1] Similarly, blacks' shorter tenure may be the result of more frequent layoffs.

Differences in educational achievement between whites and blacks may also arise because of premarket versus workplace discrimination. That is, even before blacks enter the workforce, their potential to earn a degree may be limited. If, for example, blacks receive substandard schooling at the primary and secondary levels, they may have fewer opportunities to attend college.

While differences in the explained component may in fact result from discrimination, the unexplained component may *not* reflect discrimination. This can occur if the wage model is mis-specified; that is, if an important variable that affects productivity is omitted. Economists who study male–female wage disparities frequently confront this problem. A researcher may conclude that women are discriminated against when, in fact, the gap between men's and women's earnings reflects differences in some unobservable variable, such as turnover, or other types of compensating differentials.

To understand these results better, it is useful to discuss the economics of discrimination. This gives the researcher some direction as to what factors may help disentangle the components of wage variations.

III. The Economic Theory of Discrimination

Becker (1971) provided the first analysis of the economics of discrimination. He posited that, in general, competition tends to eliminate wage differences arising from discrimination. The reason, he said, is that firms must forgo profits in order to discriminate. Any profits will be eliminated as nondiscriminatory firms enter the market to capture these windfalls. Thus, for discriminatory wage differences to exist, there can be few actual or potential nondiscriminatory firms. That is, most firms must have a preference to discriminate.

One prediction of this theory is that discriminatory wage differences will be lower in more competitive labor markets. For example, in urban areas, where there are many different job and product opportunities, white–black wage differentials resulting from discrimination should be lower than in rural areas, where fewer opportunities exist.

Even if firms prefer not to discriminate, theory predicts that they might if their customers wish to do so. If this conjecture is true, discrimination would be more prevalent in jobs with a high degree of direct customer contact. Examples of such positions include sales workers, beauty shop employees, teachers, doctors, and lawyers. Furthermore, this theory implies that racial discrimination will be lower in areas where the fraction of minority customers is large.

■ **1** Sorensen (1989) argues that women and minorities are "crowded out" of higher-paying jobs.

IV. Model Specification

Our model includes the usual independent variables that, under the null hypothesis, collectively control for individual productivity differences: female, subregions, education, potential experience (*exp*), experience squared (exp^2), experience cubed (exp^3), experience quartic (exp^4), industry (*ind*), firm size (*size*), union, occupation (*occ*), tenure (*ten*), tenure squared (ten^2), tenure cubed (ten^3), tenure quartic (ten^4), metropolitan residence (*MSA*), and marital status (*ms*).[2]

We measure educational attainment by the highest level of schooling completed: high school dropout, high school/GED, some college/vocational training, college, postgraduate (M.A.), professional, and doctorate. Because of limitations in the Current Population Survey (CPS) data, information on actual work experience is not available. Therefore, we use potential experience, measured as age minus schooling minus six, as an approximation.[3] Murphy and Welch (1990) argue that the traditional model, which includes only experience and its square, is mis-specified and that a quartic specification fits the earnings equation best.

To analyze the wage implications of industries, we define five groups: miscellaneous services (*ind* 1); manufacturing, transportation, communications, and public utilities (*ind* 2); hospital and medical services (*ind* 3); retail trade (*ind* 4); and finance, insurance, and real estate, or FIRE (*ind* 5).[4] To assess the wage effects of occupation, we define six other groups: executive, administrative, and managerial (*occ*1); professional specialty (*occ*2); sales workers (*occ* 3); service workers (*occ* 4); machine operators, assemblers, and inspectors (*occ*5); and transportation and material moving (*occ* 6).[5]

We also include a firm size variable representing companies with 1,000 or more employees. Previous studies indicate that large companies pay substantially more than their smaller counterparts. One possible reason is that large companies attract more productive workers. Other research has shown that large companies offer efficiency wages to maintain their highly skilled workforce and to reduce worker turnover (Salop [1979]; Bulow and Summers [1986])[6] or that they pay higher wages to avoid unionization (Brown, Hamilton, and Medoff [1990]). Union status itself, a potentially important variable, is measured by coverage by a union contract.

Tenure with one's current employer is also included, since human capital theory suggests that earnings differences can be accounted for by the types and levels of investment in formal or on-the-job training. As was the case for experience, we assume a quartic specification for this variable.

MSA serves as a proxy for urbanization, providing a measure of regional size and competition. The demographic variable "marital status" is equal to one if an individual has ever been married. All else equal, married females tend to earn less than their single counterparts, while the opposite holds true for males (Berndt [1991]). The interaction between female and marriage measures the differential impact that marriage has on women's wages.

The explanatory variables already mentioned are included in the model below. We have chosen earnings as the dependent variable because policy debate has focused specifically on this factor.[7]

$$X = (Subregions, Education, Experience, Experience^2, Experience^3, Experience^4, Industry, Occupation, Firm\ Size, Union, Tenure, Tenure^2, Tenure^3, Tenure^4, MSA, Female, Marital\ Status, Female \times Marriage);$$

and β is the corresponding vector of coefficients.[8]

$$(6) \quad \ln(Earnings) = X\beta + \varepsilon.$$

■ **2** Omitted conditions for all variables are as follows: female: male; education: high school dropout; industry: mining, construction, wholesale trade, private household, forestry and fisheries, public administration, educational services, social services, and other professional services; firm size: fewer than 1,000 workers; union: not covered by a union contract or no response; occupation: administrative support, including clerical, technical, and related support, handlers, equipment cleaners, and farming, forestry, and fishing; MSA: not a metropolitan area; and marital status: never married.

■ **3** For a complete discussion of the CPS measurements and their limitations, see U.S. Bureau of Labor Statistics (BLS), *Employment and Earnings,* any year.

■ **4** The use of industry dummies at the one-digit rather than the two- or three-digit level may introduce an aggregate bias.

■ **5** Listings of detailed occupations used in the CPS are published in *Employment and Earnings,* January 1994, for occupations with 50,000 or more workers. Other listings (unpublished data) can also be obtained from the BLS.

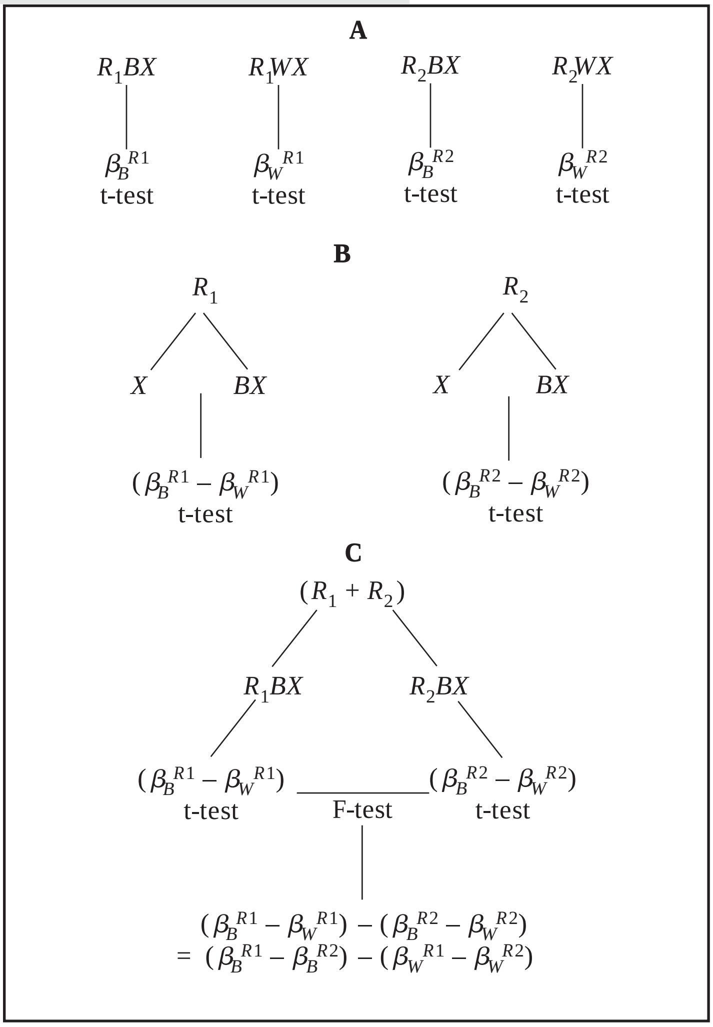
■ **6** See Katz (1986) for an overview of the literature.

■ **7** Similar results were obtained when the decomposition was done using both earnings and earnings divided by hours worked.

■ **8** The intercept term is excluded, since each region estimated includes all of its subregions.

FIGURE 1

Regional Decomposition Methodology for Blacks



SOURCE: Authors.

Data and Methodology

Different, but equivalent, regressions can be estimated to calculate the coefficients necessary for a Oaxaca decomposition. For example, including interactions between race and the independent variables is equivalent to estimating separate wage regressions for whites and blacks. Figure 1 illustrates these alternatives as well as the basic methodology used in this research.

To estimate our model, we pool regions, whites, and blacks into a single regression with interactions between region, race, and the explanatory variables (see figure 1, part C). This provides a straightforward way to determine the β's for the Oaxaca decomposition and to test whether differences in these estimates, across both regions and races, are significant.[9]

The estimates that follow are obtained by using cross-sectional microdata from the Public Use Sample of the April 1993 CPS, conducted by the Census Bureau. Wage equations for whites and blacks are estimated separately for the South and the rest of the country.[10] The sample contains full-time (35 or more hours per week), private, nonagricultural employees between the ages of 25 and 54.[11] The focus on year-round, full-time workers "minimizes earnings fluctuations due to business cycles" and "income sources other than earnings" (Levy and Murnane [1992]).

V. U.S. Wage Decomposition

Before estimating decompositions for the South and non-South, we first estimate a standard U.S. decomposition.[12] This will increase our understanding of wage differentials at the national level and will provide a reference for the regional decompositions. Table 1 presents the average wage gap between whites and blacks, breaking it down into its explained and unexplained components. The total wage differential (using whites as base) is estimated to be 32.7 percent, while the unexplained component accounts for 12.9 percent.

The explained component (variations in the means) accounts for 19.8 percentage points of the total differential. One contributing factor is education, which accounts for nearly 4 percentage points of the wage disparity between whites and blacks.

The single most important factor contributing to the explained portion of the white–black wage differential is occupation. The reason is the lower concentration of blacks in higher-

9 Although parameter estimates are identical among the three regressions, standard errors may vary. In this study, differences in standard errors are minimal.

10 The South, as defined by the U.S. Census Bureau, includes three divisions: the South Atlantic (Delaware, Maryland, the District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, and Florida), the East South Central (Kentucky, Tennessee, Alabama, and Mississippi), and the West South Central (Arkansas, Louisiana, Oklahoma, and Texas).

11 Individuals earning an average hourly wage of less than \$1 or more than \$60 are considered outliers and are eliminated from the sample (see Anderson and Shapiro [1996]). The age restriction is an attempt to capture individuals who have completed their education but have not yet retired (those in their career stages).

12 Regional dummies include all the major U.S. regions: South, Northeast, Midwest, and West.

TABLE 1

Decomposition of Wage Differentials: U.S.

Variable	Percent explained: Whites as base	Percent explained: Blacks as base	Variable (Race ×)	Percent unexplained: Whites as base	Percent unexplained: Blacks as base
	Total wage difference = 32.7%				
All	19.8 (0.0001)	20.7 (0.0001)	All	12.9 (0.0001)	12.0 (0.0001)
Education	3.9 (0.0001)	3.9 (0.0001)	Ind1	2.4 (0.001)	2.1 (0.001)
Occupation	5.7 (0.0001)	5.6 (0.0001)	MSA	2.6 (0.08)	4.7 (0.08)
Tenure	1.5 (0.0001)	1.8 (0.0001)			
MSA	3.0 (0.0001)	0.9 (0.0001)			
Marital status	2.0 (0.0001)	0.1 (0.0001)			

NOTE: Variables presented contribute most to explaining the total wage difference. P-values are given in parentheses.
SOURCE: Authors' calculations.

TABLE 2

Decomposition of Wage Differentials: South

Variable	Percent explained	Variable (Race ×)	Percent unexplained
	Total wage difference = 34.2%		
All	19.0 (0.0001)	All	15.2 (0.0001)
Education	5.0 (0.0001)	Ind1	1.6 (0.07)
Occupation	6.3 (0.0001)	Occ2	–0.7 (0.66)
Tenure	0.4 (0.0001)	Firm size	–9.5 (0.09)
MSA	3.4 (0.0001)	Tenure	–9.5 (0.43)
Female	1.6 (0.11)	MSA	2.8 (0.10)
Marital status	2.4 (0.004)	Female	5.5 (0.48)

NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

paying jobs. Indeed, executive, administrative, and managerial positions (*occ*1) account for almost half (2.8 percentage points) of the total explained differential arising from occupation (5.7 percentage points). This is the highest-paid group, with workers earning over 6 percent more than those in the professional specialty category. Only 12 percent of blacks hold managerial–administrative positions, compared to more than 22 percent of whites.

MSA accounts for another 3.0 percentage points of the explained differential (using whites as base). On average, whites earn 14 percent more in urban areas than in rural communities. At the same time, 48 percent of employed whites and 27 percent of employed blacks live in urban areas. Assuming a black wage structure, however, the contribution of *MSA* to the explained component drops to only 0.9 percent. This occurs because blacks in urban areas earn only 4.5 percent more than their rural counterparts.

Nationwide, only *MSA* and miscellaneous services (*ind* 1) contribute significantly to the unexplained portion of the wage gap. The 9.7 percent less earned by blacks relative to whites in urban areas is the reason *MSA*'s contribution to the explained component depends on the assumed wage structure (white versus black) and also explains why *MSA* is an important factor in the unexplained component. Its 2.6-percentage-point contribution to the 12.9 percent total unexplained wage differential is surprising, since one expects metropolitan areas to provide more competition, which in turn should discourage discriminatory hiring and work practices and support more equal wages.

Miscellaneous services (*ind* 1) accounts for 2.4 percentage points of the nearly 13 percent unexplained wage difference between the races. Although its total contribution is not particularly large, blacks in this category earn nearly 34 percent less than whites. This sizable gap reflects the small share of blacks (7 percent) working in this industry.

VI. Regional Variations in Earnings Inequality

The sources of white–black wage differences are not the same inside the South as outside. Tables 2 and 3 show decompositions for both regions.[13] The total wage difference between

13 Interesting differences may exist outside the South, but the small number of nonsouthern blacks in the CPS sample makes further disaggregation inappropriate.

TABLE 3

Decomposition of Wage Differentials: Non-South

	Total wage difference = 25.2%		
Variable	Percent explained	Variable (Race ×)	Percent unexplained
All	15.8 (0.0001)	All	9.4 (0.01)
Education	3.8 (0.0001)	Ind1	2.5 (0.09)
Occupation	2.3 (0.0001)	Occ2	3.9 (0.02)
Tenure	4.6 (0.0001)	Firm size	–2.8 (0.68)
MSA	1.5 (0.0001)	Tenure	–22.5 (0.07)
Female	2.3 (0.002)	MSA	2.8 (0.38)
Marital status	1.8 (0.0001)	Female	20.1 (0.07)

NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

TABLE 4

Representation of Whites and Blacks in Sample (percent)

	South		Non-South	
Variable	Whites	Blacks	Whites	Blacks
Education				
High school or less	37.6	50.2	34.8	28.9
Some college	28.6	23.0	29.7	49.3
College grad or higher	33.8	26.8	35.5	21.8
Occupation				
Highest paying	19.3	8.7	23.4	19.0
Lowest paying	6.8	17.4	5.2	2.8
Tenure (mean years)	(9.2)	(9.1)	(8.7)	(7.5)
MSA	44.7	21.9	49.7	36.6
Female	45.0	60.0	44.5	62.7
Marital status	86.0	69.4	84.4	71.1

NOTE: Variables presented contribute most to explaining the total wage difference. Mean years of tenure are given in parentheses.
SOURCE: Authors' calculations.

whites and blacks averages 34.2 percent in the South, compared to 25.2 percent outside. Nearly half of this difference can be traced to the explained components. In the South, 19.0 percent of the earnings disparity is explained, while in the non-South that share drops to 15.8 percent.

Reporting just the sum of the explained components, however, masks important regional differences. For example, occupation contributes 6.3 percentage points to the explained component in the South, but only 2.3 percentage points in the non-South. This situation is reversed for tenure, which contributes 0.4 percentage point in the South but 4.6 percentage points outside the South. These patterns occur because of important differences in the representation of blacks in certain occupations and in the average length of tenure between whites and blacks in the two regions.

Table 4 illustrates these disparities. Southern blacks are underrepresented in the highest-paying occupation (administrative–managerial) and disproportionately represented in the lowest-paying one (operatives). These differences are virtually nonexistent in the rest of the country. On the other hand, tenure levels in the South are nearly identical between the races, while in the non-South blacks average over a year less on their current jobs.

Tables 2 and 3 indicate that educational differences between the races explain 5.0 percentage points of the white–black wage gap in the South and 3.8 percentage points outside the South. Despite the similarity of these numbers, table 4 shows that blacks have strikingly different patterns of educational attainment in the South versus the non-South. These differences are not captured if one considers just the mean years of education for blacks in the two regions (13.3 years in the South, 13.5 years in the non-South).

Substantially more blacks attend college outside the South than in the South—71 percent versus 50 percent. Interestingly, though, more blacks in the South obtain their four-year college degree (26.8 percent, compared to 21.8 percent outside the South). This occurs because 54 percent of blacks who attend college in the South obtain at least a bachelor's degree, while only 31 percent of blacks outside the South do. If blacks in the non-South completed college at the same rate as their counterparts in the South, our model indicates that their wages would increase by 2.3 percentage points, narrowing the white–black earnings gap to 22.9 percent. The share of the white–black wage differential that is unexplained is 15.2 percent in the South and 9.4 percent in the non-South—both highly significant.

TABLE 5

Regional Parameters and Differences

Variable (Race ×)	Parameter estimates South	Non-South	Difference
Ind 1	−0.25 (0.07)	−0.30 (0.09)	0.05 (0.84)
Occ 2	0.06 (0.66)	−0.37 (0.02)	0.43 (0.04)
Firm size	0.12 (0.09)	0.04 (0.68)	0.08 (0.49)
MSA	−0.13 (0.10)	−0.077 (0.38)	−0.063 (0.66)
Female	−0.09 (0.48)	−0.32 (0.07)	0.23 (0.30)

NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

TABLE 6

Unexplained Wage Differences for Various Groups

Variable	South	Non-South	Difference
All	15.2 (0.0001)	9.4 (0.01)	5.8 (0.24)
All – Occ 2	15.9 (0.0001)	5.6 (0.18)	10.3 (0.06)
All – Ind 2	13.6 (0.0002)	6.9 (0.09)	6.7 (0.20)
Males	20.1 (0.0001)	6.5 (0.28)	13.6 (0.08)
Females	11.9 (0.004)	11.2 (0.02)	0.7 (0.90)
Urban	22.1 (0.0005)	10.3 (0.09)	11.8 (0.18)
Rural	13.1 (0.0003)	4.2 (0.73)	8.9 (0.51)
Large firms	10.4 (0.17)	7.0 (0.12)	3.4 (0.70)
Small firms	25.7 (0.0001)	15.6 (0.03)	10.1 (0.28)

NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

Unexplained Wage Differentials

Tables 2 and 3 also report the breakdown for the independent variables in which unexplained wage differences are significant in one or both regions. Only miscellaneous services (*ind* 1), which was important for the U.S. as a whole, is significant in both the South and non-South, accounting for 1.6 and 2.5 percentage points of the unexplained differences, respectively. Within that occupational category, blacks earn 25 percent less than their white counterparts in the South and 30 percent less than whites outside the South (see table 5).

The biggest difference between the two regions occurs within the professional specialty category (*occ* 2). In the South, *occ* 2 does not account for any of the unexplained component of the white–black wage differential, whereas in the non-South it accounts for 3.9 percentage points of the 9.4 percent unexplained wage gap. This is because within this occupation blacks outside the South earn an average of 37 percent less than whites (see table 5). In the South, the sign is reversed but insignificant.

The difference in the way whites and blacks are rewarded within professional specialty occupations is highly significant (p-value = 0.04) and impacts the two regions' race-based pay gaps. Excluding this category, the unexplained earnings disparity in the non-South drops to 5.6 percent and becomes insignificant. This is in sharp contrast to the 15.9 percent unexplained pay differential seen in the South when this group is excluded (see table 6).

In the South, the variable with the greatest impact on the unexplained pay component is firm size. Whites working in large firms earn 3.9 percent more than those working in small firms, but the earnings differential for blacks is much greater—16.3 percent. The pay difference between whites and blacks in large firms is 10.4 percent and is insignificant (see table 6); in small firms, that share jumps to 25.7 percent.

One commonly held belief is that the narrowing of the white–black wage gap between 1960 and 1980 is attributable to antidiscrimination laws and the genesis of affirmative action programs. The substantial premium associated with firm size in the South could occur if large firms there are policed much more closely than small firms or if the laws are enforced unequally.

MSA is significant only in the South, but contributes 2.8 percent to the total unexplained wage differential in both regions (see tables 2 and 3). This occurs because a much greater share of working blacks outside the South live

TABLE 7

Non-South Parameter Estimates for Black × Tenure and × Female

Original: Variable (Black ×)	Original: Parameter estimates	Variable (Black × male)	Parameter estimates	Variable (Black × female)	Parameter estimates
All tenure [a]	−0.225 (0.07)	All tenure [a]	−0.0153 (0.80)	All tenure [a]	−0.3431 (0.04)
Ten	0.1246 (0.10)	Ten	0.0015 (0.99)	Ten	0.1969 (0.04)
Ten^2	−0.0155 (0.14)	Ten^2	−0.0034 (0.88)	Ten^2	−0.0226 (0.09)
Ten^3	0.0007 (0.20)	Ten^3	0.0003 (0.81)	Ten^3	0.00087 (0.19)
Ten^4	−0.000009 (0.27)	Ten^4	−0.000006 (0.78)	Ten^4	−0.0000106 (0.30)
Female	−0.32 (0.07)			Black × female	−0.672 (0.02)

a. This estimate is the unexplained component for tenure.
NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

in urban areas (36.6 percent versus 21.9 percent; see table 4). The importance of *MSA* in the South may arise because blacks make up more than 50 percent of the rural population there, but only 22 percent of the urban population.

The next section takes a closer look at the reasons behind these regional differences. It also attempts to ascertain whether some of the explained and unexplained differences are due to productivity and discrimination, respectively.

Regional Differences: A Closer Look

By analyzing the apparent differences between the two regions, this section illustrates why teasing out discrimination is so difficult. Tables 2 and 3 show that in the non-South, blacks gain significantly more than whites from an additional year on the job (p-value = 0.07), and black women earn 32 percent less than comparable black men. In the South, there is no significant difference for either variable. Although our strict interpretation of the null is that discrimination in the non-South is associated with black women only, there are alternative explanations.

Table 4 shows that the non-South is also the region with important tenure differences between the races. In the South, tenure levels for whites and blacks are almost identical, with no differential return. This suggests that certain unobservable characteristics may cause higher turnover and hence lower wages for blacks in the non-South. As blacks accumulate time on the job, the importance of this unobservable diminishes.

Evidence suggests that, on average, women have weaker attachments to the labor force than do men and as a result earn less. Firms, however, probably cannot tell at the time of hire which women are more likely to stay. Since turnover is costly, firms must be compensated whenever they hire someone they believe will have a greater chance of leaving. A woman may initially start out at a lower salary than a man doing the same job, but should correspondingly gain more per additional year of tenure. Women should then be compensated as they slowly reveal themselves to be "stayers."

The large return to blacks from tenure, coupled with the large negative return to black women, suggests a connection between the two. For example, black women may have a weaker attachment to the labor force than do black men, and this difference may be greater than it is for whites. To investigate whether this explanation can account for the unusually large returns to tenure and the small returns to black women, we break the black-tenure interaction into its male and female components. The

TABLE 8

South Parameter Estimates for Black × Tenure and × Female

Original					
Variable (Black ×)	Parameter estimates	Variable (Black × male)	Parameter estimates	Variable (Black × female)	Parameter estimates
All tenure [a]	0.095 (0.43)	All tenure [a]	–0.473 (0.02)	All tenure [a]	0.147 (0.46)
Ten	–0.023 (0.73)	Ten	–0.187 (0.06)	Ten	0.272 (0.05)
Ten^2	0.002 (0.78)	Ten^2	–0.022 (0.08)	Ten^2	–0.033 (0.10)
Ten^3	–0.0001 (0.78)	Ten^3	0.001 (0.09)	Ten^3	0.0014 (0.18)
Ten^4	–0.000002 (0.80)	Ten^4	–0.00001 (0.11)	Ten^4	–0.000002 (0.31)
Female	–0.092 (0.48)			Black × female	–0.742 (0.03)

a. This estimate is the unexplained component for tenure.
NOTE: Variables presented contribute most to explaining the total wage difference (whites as base). P-values are given in parentheses.
SOURCE: Authors' calculations.

expectation is that black tenure will be significant for women and insignificant for men.

Table 7 confirms this suspicion, showing that black females gain more than whites for extra years on the job, while the returns to tenure for black males and whites are nearly equal.[14] In other words, tenure helps equalize the wage difference between white and black females but not between white and black males. As a result, the black female coefficient falls even further and becomes more significant. Before interactions with tenure, black women earn 32 percent less than white women; afterward, they earn 67 percent less. This is because of the 34 percent that tenure adds to the wages of black females compared to whites. The average wage difference is about 33 percent, close to the 32 percent estimated before allowing interactions. Because the importance of tenure and female in the non-South could be due to model misspecification, the unexplained differences may not reflect labor market discrimination.

Table 8 shows the same experiment for the South. Unlike earlier (table 5), the black female coefficient is now highly significant and nearly identical to that of the non-South. Therefore, the surprising finding that black women earn less than black men in the non-South could reasonably be attributed to model mis-specification. Tables 7 and 8 reveal that the major difference between the two regions with regard to tenure is that tenure offers a larger return to blacks in the non-South than in the South. This may reflect the greater tenure differences between whites and blacks in the non-South.

If the return to black females is indeed due to model misspecification, then perhaps the unexplained returns to black males is a better indicator of potential discrimination. If these differentials are interpreted as reflecting discrimination, we would conclude that discrimination is important in the South (the unexplained component is 20.1 percent and is highly significant [p-value = 0.0001]) but not outside the South. If this interpretation is correct, there is little evidence to suggest that non-South discrimination is important, since the unexplained return is only 6.5 percent and is insignificant.

The single most important factor contributing to the explained wage differences between whites and blacks is occupation—6.3 percent in the South and 2.3 percent outside the South. Although these differences are "explained," as mentioned before, they may reflect discrimination if blacks are segregated into lower-paying occupations. Probit estimates indicate that educational differences between whites and blacks in the non-South can explain why 19 percent of blacks are in the highest-paying occupations

14 We also tried an interaction between female and tenure for whites, but it was insignificant; therefore, we use whites as the reference group.

TABLE 9

Non-South Professional Specialty Occupation

Variable (Race ×)	Parameter estimates
Public	–0.635
	(0.006)
Nonpublic	–0.182
	(0.345)

NOTE: P-values are given in parentheses (whites as base).
SOURCE: Authors' calculations.

(*occ* 1), compared to 23.4 percent of whites (*p*-value = 0.49). This is not the case in the South. Even after controlling for educational differences, blacks are significantly less likely than whites to work in *occ* 1 (*p*-value = 0.0002). This provides some ammunition for the argument that blacks are more likely to be segregated into lower-paying jobs in the South than in other areas of the country.

Another argument is that differences in white–black tenure levels in the non-South may reflect discrimination (that is, the 4.6-percentage-point "explained" tenure difference is due to discrimination). This could arise if blacks in the non-South are fired "without cause" more often than are whites. Basically, the question is whether wages between whites and blacks adjust more freely in the South, whereas quantities adjust more freely in the non-South. It is impossible to determine conclusively whether tenure differences in the non-South result from discrimination; however, the following argument casts doubt on discrimination's ability to explain the lower tenure levels for blacks in the non-South.

If blacks are being laid off without cause more frequently in the non-South than in the South, then one would expect the difference between potential experience (which is observable) and actual experience (which is not observable) to be greater in the non-South. This is because unemployment spells are much longer for laid-off workers than for those who quit. If blacks in the non-South are indeed more likely to be laid off, then they should also have more interruptions in their work histories. Thus, the ratio of actual experience to potential experience for blacks in the non-South would be lower, or equivalently, the measured returns to potential experience would be lower. No such difference should exist in the South, however. The data show that there is no significant difference between the returns to potential experience for whites and blacks in either region; furthermore, the sign is the opposite of that expected. This casts doubt on the proposition that discrimination is reflected in wage differences in the South and quantity differences outside the South.

Another important difference between the regions is in how whites and blacks are compensated within professional specialty occupations. Outside the South, whites within this category earn significantly more than blacks; in the South, however, no such difference exists. Faced with this observation, one might appeal to discrimination theory and speculate that the source of the effect is customer-driven discrimination. The next regression addresses this possibility by examining whether white–black wage differences within professional specialty occupations outside the South arise because the category comprises public jobs. (See the appendix for groupings of public and nonpublic jobs.)

Rerunning the equation for the non-South suggests important differences in how blacks are compensated within this category. In public occupations (those with more customer contact), blacks earn 63.5 percent less than their white counterparts (*p*-value = 0.0056); as table 9 shows, this difference is insignificant in nonpublic occupations (*p*-value = 0.35). The difference between these two figures is only marginally significant (*p*-value = 0.11).

VII. Concluding Thoughts

There is little evidence to suggest that white–black wage disparities outside the South cannot be explained by differences in observable characteristics. The situation is much less clear in the South, where the significance of the unexplained components of the race-based pay gap is robust. Furthermore, even some of the "explained" differences in the South suggest the possibility of workplace discrimination. For example, unlike the rest of the nation, the South cannot attribute blacks' underrepresentation in high-paying occupations to differences in observed characteristics like education.

The other major factors driving the lower wages of blacks in the South are firm size and location: Blacks working in small firms or urban areas earn significantly less than comparable whites. It is difficult to determine whether discrimination is behind either of these unexplained wage differences. Although the lesser amount earned by urban blacks is consistent with discrimination, it may also arise because of unobservable differences between rural versus urban blacks. The difficulties heterogeneity causes in trying to measure discrimination cannot be overemphasized; discrimination exists for both whites and blacks, even after controlling for observable characteristics. And evidence suggests that this problem is getting worse. Income inequality over the last 30 years has increased for both whites and blacks; however, it has been especially pronounced for blacks. Understandably, identifying the true source of these trends and differentials represents a major challenge for economists and policymakers alike.

Appendix:

Public and Nonpublic Occupational Groupings

Miscellaneous Services Industry

Public
- Hotels and motels
- Lodging places, except hotels and motels
- Laundry, cleaning, and garment services
- Beauty shops
- Barber shops
- Funeral services and crematories

Nonpublic
- Advertising
- Services to dwellings and other buildings
- Personnel supply services
- Computer and data processing services
- Detective and protective services
- Business services
- Automotive rental and leasing, without drivers
- Automobile parking and car washes
- Automotive repair and related services
- Electrical repair shops
- Miscellaneous repair services
- Theaters and motion pictures
- Video tape rental
- Bowling centers
- Miscellaneous entertainment and recreation services

Professional Specialty Occupation

Public
- Health diagnosing occupations
- Health assessment and treating occupations
- Teachers, college and university
- Teachers, except college and university
- Counselors, educational and vocational
- Librarians, archivists, and curators
- Social scientists and urban planners
- Social, recreation, and religious workers
- Lawyers and judges

Nonpublic
- Engineers, architects, and surveyors
- Mathematical and computer scientists
- Natural scientists
- Writers, artists, entertainers, and athletes

References

Anderson, Deborah, and David Shapiro. "Racial Differences in Access to High-Paying Jobs and the Wage Gap between Black and White Women," *Industrial and Labor Relations Review*, vol. 49, no. 2 (January 1996), pp. 273–86.

Becker, Gary S. *The Economics of Discrimination*, 2d ed. Chicago: University of Chicago Press, 1971.

Berndt, Ernst R. *The Practice of Econometrics: Classic and Contemporary*. Reading, Mass.: Addison–Wesley, 1991, pp. 150–223.

Brown, Charles, James Hamilton, and James Medoff. *Employers Large and Small*. Cambridge, Mass.: Harvard University Press, 1990.

Bulow, Jeremy I., and Lawrence H. Summers. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment," *Journal of Labor Economics*, vol. 4, no. 3, part 1 (July 1986), pp. 376–414.

Katz, Larry. "Efficiency Wage Theories: A Partial Evaluation," in Stanley Fischer, ed., *NBER Macroeconomics Annual*, 2d. ed. Cambridge, Mass.: MIT Press, 1986.

Levy, Frank, and Richard J. Murnane. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, vol. 30, no. 3 (September 1992), pp. 1333–81.

Murphy, Kevin M., and Finis Welch. "Empirical Age–Earnings Profiles," *Journal of Labor Economics*, vol. 8, no. 2 (April 1990), pp. 202–29.

Oaxaca, Ronald L. "Male–Female Wage Differentials in Urban Labor Markets," *International Economic Review*, vol. 14, no. 3 (October 1973), pp. 693–709.

_____, and Michael R. Ransom. "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, vol. 61, no. 1 (March 1994), pp. 5–21.

Salop, Steven. "A Model of the Natural Rate of Unemployment," *American Economic Review*, vol. 69, no. 1 (March 1979), pp. 117–25.

Sorensen, Elaine. "The Crowding Hypothesis and Comparable Worth," *Journal of Human Resources*, vol. 25, no. 1 (February 1989), pp. 55–89.