THE 1990 CENSUS AND STATISTICAL ADJUSTMENT

by

David A. Freedman

Technical Report No. 279
January 1991

Department of Statistics
University of California
Berkeley, California

THE 1990 CENSUS AND STATISTICAL ADJUSTMENT

by

David A. Freedman

Abstract

Considering the difficulties, the Census does a remarkably good job at counting people. This article outlines the process, and reviews the two current techniques for evaluating or adjusting the Census. Demographic analysis uses administrative records to make independent population estimates, which can be compared to Census counts. Capture-recapture uses data from an independent sample survey to estimate population coverage in the Census. These techniques do not seem solid enough to be used for adjustment; indeed, they could easily make the counts less accurate.

Author's footnote. D. A. Freedman is Professor of Statistics, University of California, Berkeley, Ca 94720.

THE 1990 CENSUS AND STATISTICAL ADJUSTMENT

An Overview of the Census

The Census has been taken every ten years since 1790. It is a sophisticated enterprise whose scale is remarkable. There are about 9,000 permanent staff. Between October 1989 and September 1990, these staff open 500 field offices, in which they hire and train 500,000 temporary employees. In Spring 1990, a Census media campaign encourages people to cooperate with the Census and asks "Were You Counted?"-- in English, Spanish, and several Asian languages. (Most of the resources are donated by public broadcast media.) The Census is ahead of Coca Cola, at least with respect to volume of advertising (1).

The population of the United States in 1990 is about 250 million persons in 100 million households, distributed across 7.5 million "blocks" (the smallest units of Census geography). Statistics for larger areas like tracts, cities, or states are obtained by adding up component blocks (2). Most of the population lives in TAR (Tape Address Register) areas, mainly urban. The Census begins by acquiring and merging commercial address lists. Each address can be assigned to a census block ("geocoded") using computer files whose development started in 1980. Address lists are checked for completeness by the Post Office and in a "pre-canvass" by Census personnel who walk the streets and look for omitted housing units. In pre-list areas (typically suburban), the address list is developed by Census personnel who canvass the area, making spot maps to show the location of each residence, which is then geocoded (3).

In TAR and pre-list areas, Census forms are mailed to each household, filled out by a respondent and mailed back (4). Nationally, the mail-back rate was 78% in 1970 (when the mail-out-mail-back procedure was used in TAR areas), 75% in 1980, and 63% in 1990 (5). In 1990, for the first time, computers are used to log forms in and out, and to organize "non-response follow-up" (6). Households that do not return forms are followed up by enumerators: 3 visits (and 3 telephones calls) are made, and then "last resort" information may be obtained from neighbors, building superintendents, etc. After follow-up is complete, Census response rates approach 100%. Forms are microfilmed, scanned by optical readers (10,000 forms an hour), and checked for incomplete or inconsistent entries. Errors trigger follow-up by telephone or personal visit (7).

The Census provides a statistical portrait of the United States, at ten-year intervals. Geographical detail makes these data unique. Fortunately or unfortunately, the counts have more than academic interest: they influence the distribution of power and money. The Census is used to apportion Congress as well as local legislatures, and to allocate tax money-- $40 billion per year in the late 1980s-- to 39,000 MCDs (Minor Civil Divisions) (8,9). For these purposes, the geographical distribution of the population matters, rather than counts for the nation as a whole. Indeed, the Census is used as a basis for sharing out fixed resources: if one jurisdiction gets more, another must receive less. In this context, adjusting the Census is advisable only if the process brings us closer to a true picture of the distribution of the population-- a tall order.

The Undercount

The Census does a remarkably good job at counting people-- with occasional well-publicized lapses (10). However, it is believed that a small undercount remains. This undercount has many causes, and is not uniform. People who move at Census time are hard to count. In rural areas, maps and address lists are incomplete. Central cities have heavy concentrations of poor and minority persons, who may be less cooperative with government agencies.

If the undercount can be estimated with good accuracy, especially at the local level, adjustments can be made to improve the Census. Some statisticians believe that the undercount can be estimated well enough, others are skeptical: a bad adjustment may be worse than nothing (11). How can the undercount be estimated? One direct method is to take a sample of small areas, and count them more accurately. Census counts could then be calibrated, by comparison. However, current methods for estimating the undercount do not work that way, and are quite indirect. The two methods under consideration for 1990 are demographic analysis and capture-recapture.

In outline, demographic analysis relies on vital statistics (births and deaths) and other administrative records to make independent estimates of the national population. Capture-recapture uses survey data collected after the Census to estimate coverage at the local level. If the undercount is large, these estimates may be accurate enough for adjustment. With a small undercount, it is unlikely that current adjustment methodologies can improve on the Census; instead, adjustment could easily degrade the accuracy of the data.

Because of its resource implications, the undercount has attracted considerable attention-- in the media, the Congress, and the courts. After the 1980 Census, New York City (and other jurisdictions) sued the Department of Commerce, seeking to compel an adjustment based on demographic analysis and capture-recapture. The Department resisted this pressure, and was upheld by the Court, which found "as a matter of fact that a statistical adjustment of the 1980 census is not feasible" (12).

For 1990, the Department of Commerce again decided not to adjust the Census, and was again sued by New York City and some of the other 1980 plaintiffs; the issues in the two cases seem quite similar. One part of the 1990 suit was settled before trial: the Secretary of Commerce agreed to reconsider, and make a new decision on adjustment by July 15, 1991 (13).

Demographic Analysis

Demographic analysis makes independent estimates of the national population using administrative records. The starting point is an accounting identity:

Population = Births - Deaths + Immigration - Emigration

The estimates are made by age, sex and race (white, black, other), and compared to Census counts. For example, the undercount in 1970 was estimated at about 3% nationally; in 1980, about 1%. (The 1990 estimate is likely to fall between 1% and 3%.) The undercount for blacks was estimated as about 5 percentage points above whites, in 1970 and 1980 (14).

Estimates cannot be made for the hispanic population because vital statistics lack information on ethnicity (15). Data on emigration are incomplete. And there is substantial illegal immigration, which cannot be measured directly. In 1980, for instance, it is estimated that roughly 3 million illegal immigrants were living in the U.S.; about 2 million are thought to have been counted in the Census (16).

Evidence on differential undercounts depends on racial classifications, which may be problematic; and procedures vary widely from one data collection system to another. For the Census, race of all household members is reported by the person who fills out the form. On death certificates, race of decedent is often determined by the undertaker. Birth certificates show the race of the mother and (usually) the race of father; procedures for ascertaining race differ from hospital to hospital. Birth certificate data are collected by the National Center for Health Statistics, which uses a computer algorithm to determine race of infant from race of parents. The algorithm maximizes the chance that an infant will be classified as non-white. Changing the algorithm is likely to reduce recent estimated undercount rates for young black children by 2-5 percentage points (17).

Coverage of vital statistics is another problem, and demographic analysis uses different techniques for different age groups, with further variations by race and sex. In the period 1935-1960, the coverage of the birth certificate system was far from complete, especially for blacks. Thus, in order to estimate undercount rates for persons age 30-55 in 1990, birth certificate data must be adjusted for under-registration; and the adjustment is based on Census data. In short, before birth certificate data can be used to adjust the Census, the Census must be used to adjust the birth certificates (18).

Prior to 1935, many states did not have birth certificate data at all; and the further back in time, the less complete is the system. This makes it harder to estimate the population age 55 and over. For the period 1925-35, a set of data for whites was created by Whelpton (19); he estimated what the data would have been if the states had been in the birth certificate system; his data are even adjusted for under-registration that would have occurred. For black females, another set of imputed data is used, and the number of black males is estimated using expected sex ratios. There is yet another set of imputations for persons of other races (e.g., Asians), completing the analysis for persons age 55-65 in 1990. Finally, persons born before 1925 will be over 65, and eligible for Medicare. Demographic analysis estimates the number of such persons starting from Medicare records (adjusted for under-enrollment) rather than birth records.

Figure 1. Results from demographic analysis: Estimated undercount rate for black males, by age group, in the Census of 1960, 1970, and 1980 (20).

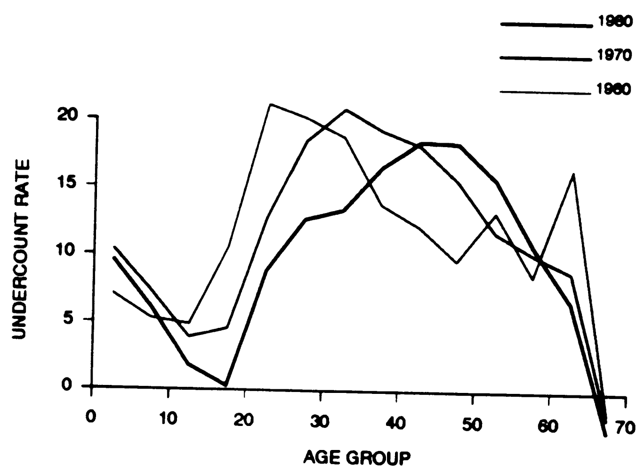

Figure 1 shows results from demographic analysis-- the estimated undercount rates by age for black males in three Census years (1960, 1970, 1980). The shape of the curves suggests a failure in the logic of demographic analysis. In 1980, for instance, the undercount is about 10% for new-borns, dropping nearly to 0% in the late teens, rising to 18% at age 40-44, then dropping again. A more plausible pattern is just the reverse: high undercount rates for teenagers and low rates for the middle-aged.

Moreover, the age group with the highest undercount rate was 20-24 in 1960, 30-34 in 1970, and 40-44 in 1980. Now there may be a cohort of hard-to-count people who were aged 20-24 in 1960, and who remained hard to count as they aged through the Census of 1970 and 1980. A more plausible explanation is statistical artifact in the demographic analysis-- including over-adjustment of black births in the 1930s (22).

The validity of demographic analysis depends on a series of complicated adjustments to a variety of administrative statistics. The errors may be small, but so is the undercount. To make demographic analysis useful for adjustment, the errors have to be quantified with some precision. This does not seem feasible (23).

One limitation of demographic analysis is widely recognized. The estimates are national rather than local, because data are lacking on internal migration. Of course, national undercount rates can be applied to small areas, a process called "synthetic estimation." For instance, if the undercount rate for black males age 40-44 is estimated as 18% for the country as a whole, the number of such persons in every block can be increased by the "adjustment factor" 1/(1-.18)= 1.22. This method is of doubtful utility for making small-area estimates, since undercount rates must vary by substantial amounts from place to place.

Capture-Recapture: The DSE

To estimate the undercount in geographic detail, proponents of adjustment suggest using the DSE ("Dual System Estimator") on capture-recapture data. This is easiest to explain in a hypothetical example: estimating the number of fish in a lake. You catch 1,000 fish, tag them and throw them back (the "capture"). Then catch another 100 fish ("recapture"). Say that 90 of the recaptured fish are tagged, suggesting that 90/100 = 90% of the fish in the lake are tagged. Then the total population-- tagged and untagged-- would be estimated as

$$\frac{1{,}000}{90/100} = 1{,}111 \qquad (1)$$

The idea is appealing, but there are limitations: Recapture has to be done at random (equal probability of recapture for all fish), tags have to stay on the fish, and the population of the lake has to stay the same between capture and recapture (24).

For the DSE, "capture" means being counted in the Census. Recapture is by a special sample survey, the PES (Post Enumeration Survey). The PES is based on a stratified sample of about 5,000 blocks, with 150,000 households and 400,000 people (25). PES interviewers go through the sample blocks, list all the households they find, and interview one respondent in each household.

The PES is unlikely to have better coverage than the Census. For example, the Current Population Survey (done monthly by the Bureau to estimate employment statistics) seems to be sampling from a population that is 5%-10% smaller than the Census population. The completeness of the Census may derive from its community outreach and advertising campaigns. Such measures are less helpful in a sample survey like the PES, where few people know whether or not they fall into the sample.

Since the PES does not improve on the Census by counting better, the capture-recapture model is essential. In principle, it may be easy to see whether a fish is tagged. With people, the problem is more complicated. Records from the PES have to be matched against records from the Census, to determine whether a person in the PES was captured in the Census. However, record matching is a complex and error-prone process. In general, mistakes tend to create false non-matches. This reduces the denominator of the DSE, and inflates the estimated undercount: see equation 1 or 2.

The Census is taken in the spring and counts people at their usual place of residence on Census day (April 1). The PES is done in the summer, with follow-up interviews in the fall and winter (26). Matching 400,000 records from the PES against 250 million records from the Census is virtually impossible. The search area must be restricted, and this is done by address on Census day.

Determining the right Census-day address for PES respondents is a critical step, and one that may be quite hard to do. About 20% of the population moves every year, late spring and early summer being peak times. In some areas, roughly one third of the population has moved between Census day and PES interviewing; 5-10% may be more typical.

Ordinarily, the PES interviews one respondent in a household. On the basis of this interview, the PES must determine the usual place of residence for all members of that household, for a particular day some three to nine months in the past. This is a tall order-- particularly for movers and for households consisting of unrelated people living together, where the roster can be expected to change substantially over a period of several months. Failure in tracking people back to their Census day address is likely to have a serious impact on the estimates.

The Census does not collect unique identifiers like social security numbers (27); the matching algorithm uses name, address, age, sex, race and ethnicity. Of course, some of the data are inaccurate, on the PES side as well as the Census side. There are variations in spelling, and some persons give fictitious names. Demographic characteristics (even sex) sometimes appear to change from one interview to another (28).

About 70% of the matching is done by computer; the rest, by clerks. For many of the records, match status cannot be determined on the basis of the information obtained in the Census and the initial PES interview. Such cases are re-interviewed ("sent to follow-up"). Even so, some cases remain "unresolved" and statistical models are used to impute their unknown match status. The validity of these models is questionable. In 1980, roughly 8% of the PES cases were imputed. In 1990, the problem is expected to be less severe, but even a small percentage of missing data spells trouble when the undercount rate is small (29).

The DSE must deal with erroneous enumerations in the Census, including double-counting and fabrication (30). To deal with these problems, and limit the search area for matches to manageable scale, the DSE uses the "P-sample" and "E-sample". The P-sample consists of all the people (in the sample blocks) found by the PES interviewers. The E-sample consists of Census records for these same blocks.

An attempt is made to match persons in the two samples: a match validates both the Census and the PES records (31). Persons in the P-sample but not the E-sample represent "gross omissions": they may have been missed by the Census. (Of course, they may also have been counted in error by the PES, e.g., at the wrong address.) Persons in the E-sample but not the P-sample represent potential erroneous enumerations. Finally, there are persons in neither sample: their existence cannot be demonstrated directly, but their number is estimated by statistical modeling. The resulting classification is shown in a 2 x 2 table:

		E-sample (Census) In	E-sample (Census) Out
P-Sample (PES)	In	matches	gross omissions
	Out	erroneous enumerations	not found

The formula for the DSE is (14)

$$\text{DSE} = \frac{\text{Cen} - \text{EE} - \text{II}}{\text{M/N}} \qquad (2)$$

The notation is as follows:

DSE = Dual System Estimate of population.

Cen = Census count.

EE = estimated number of erroneous enumerations.

II = number of imputations and unmatchable persons in the census. (Some persons are imputed into the Census, or counted without enough detail for matching; such persons have to be subtracted out of the Census count.)

N = estimated population obtained by weighting up P-sample block counts.

M = estimated total number of matches obtained by weighting up sample matches (32).

"Weighting up" may be an unfamiliar term, but the idea is easy: if you sample 1 block in 1,000, say, then each sample person counts for 1,000 in the population--and therefore gets a "sample weight" of 1,000. Population subtotals (e.g., for matches) are estimated by adding up sample weights for the corresponding people. In practice, there are different sample weights in different strata, and adjustments are made for non-response.

Intuitively, the "match rate" M/N in the denominator of the DSE estimates the fraction of the population counted by the Census. The ratio DSE/Cen is an "adjustment factor": it adjusts the Census count to the Dual System Estimate.

Post Stratification, Smoothing, and Block-Level Adjustment

Different kinds of people are likely to have unequal probabilities of responding to the PES, violating the randomness assumption for recapture. As a partial solution, the PES sample is "post stratified" by 6 age groups, by sex, race, ethnicity, and housing "tenure" (owner or renter). There are about 1,400 post strata (33). The DSE and corresponding "raw" adjustment factor are computed separately for each one. On average, we expect 400,000/1,400 ≈ 300 sample people in each post stratum, and only a few gross omissions and erroneous enumerations. Resulting estimates would be quite unstable, due to sampling error.

To reduce sampling error, statistical smoothing techniques are used to average results from similar post strata. More technically, regression models are used to predict adjustment factors using some of the variables which define strata and post strata, and predicted factors are averaged with raw factors (34). However, bias (e.g., from matching error) remains a problem. Furthermore, statistics which measure the reliability of smoothed results can be quite misleading unless the models are valid. With the DSE, the models are questionable, and there are no external benchmarks (35).

To adjust block-level counts, synthetic estimates are proposed, starting from post strata. For example, take black and hispanic males 45-64 living in central cities in New England. Suppose the DSE for this post stratum is 10% over the Census count, so the adjustment factor is 1.1. Now suppose some central-city block in New England has a Census count of 10 black or hispanic males age 45-64. According to the DSE, there are 1.1x10=11 such persons in the block (36). One of the 10 real Census records is chosen at random and copied. The resulting fictitious person is added to a special "adjustment category" in the block, and comes into all Census of Population tables for areas that include the block. This scenario is repeated for every block, increasing the post stratum count by 10%: block-to-block variability is ignored.

Some post strata will have adjustment factors below 1.00, corresponding to apparent overcounts (37). Suppose a central-city block in New England has 20 white males age 45-64, by Census count; and the adjustment factor is .95. According to the DSE, there are only .95x20=19 such people in the block. One of the Census records is selected at random and a corresponding "negative person" is put into that block's special adjustment category. This process is applied uniformly to all blocks, reducing the post stratum count by 5%. Real people are subtracted from the Census tables.

Undercount rates vary from one geographical area to another, and from one demographic group to another. That is why synthetic estimates for small areas, based on demographic analysis, have not been widely accepted. However, the same problem affects the DSE and makes its block-level adjustment unsatisfactory. For example, one post stratum consists of hispanics -- cross classified by age, sex, and housing tenure -- in central cities in the Pacific Division (California, Washington, Oregon, Alaska and Hawaii). In round numbers, the 1990 population of the Pacific Division is likely to be 40 million with 8 million hispanics, 5 million of the latter being in southern California.

Consider an adjustment for Stockton, a city of about 200,000 people in California's Central Valley, a four hour drive north of Los Angeles. The hispanic population is about 50,000; there can be at most a few dozen hispanics from Stockton in the PES, and a handful of gross omissions or erroneous enumerations. No stable estimates could be developed from a sample that small. Instead, estimates for Stockton will be based on the adjustment factor for the whole post stratum, and the numbers will be driven by PES data from southern California. The basic assumption: undercount rates for hispanics are the same in Stockton as in Los Angeles-- or Honolulu. There is no empirical evidence to support this assumption.

The Los Angeles Test

Quantitative evaluation of the 1990 PES waits on publication of the data. However, the methodology was tried in 1986 in the Los Angeles Test of Adjustment Related Operations ("TARO"). A test Census was done in a heavily hispanic area with a population of about 300,000. Then, a test PES was done, estimating an undercount of 9%. This estimate reflects errors not only in the Census, but also in the adjustment methodology, e.g., matching errors.

In TARO, special follow-up research studies were done, and some information is available on the magnitude of these errors. The data can be interpreted in many ways, and considerable uncertainty remains (38). Still, over half the estimated undercount seems to be due to PES errors, so adjustment by the DSE would move the Census from an undercount of a few percent to a somewhat larger overcount. For example, about 5% of the data from the E-sample were imputed, with significant under-statement of erroneous enumerations. Alternative and more plausible imputations reduce the estimated undercount by 2 percentage points.

Another 3 percentage points of undercount seem to come from respondents who gave bad address information at the PES interview. Many of these respondents were in fact movers but said otherwise, creating false non-matches and inflating the estimated undercount. (Indeed, about half this group seems to have moved into the test site after the test census: such respondents are "out of scope".) There are many other sources of error. For example, fabrication by PES interviewers may contribute a percentage point to the estimated undercount.

Proponents of adjustment reply by estimating how many movers would have been correctly classified had they given correct addresses and been in scope, or how many fabricated interviews would have matched to the test census had they been real (39). Such estimates seem fanciful.

Summary and Conclusion

The Census does a remarkably good job at counting people-- given the difficulties in large-scale statistical work. Still, an undercount may be expected. Of the two current adjustment methodologies, demographic analysis must cope with small errors and inconsistencies in a variety of administrative data systems; its estimates are made only at the national level. The Dual System Estimator faces problems created by incorrect or missing data-- especially for movers-- which increase the error rate in record matching and inflate estimated undercounts. Variation in undercount rates from place to place is a reality faced by both methodologies.

There is little hard evidence to show that current adjustment methodologies would improve the accuracy of the Census, and much can go wrong. In short, the present state of the art does not support adjustment in 1990.

Footnotes for Census article

(1) Vitt Media International, 1990 Census Campaign Awareness Study.

(2) A few blocks do cross administrative boundaries (e.g., city lines); and are "split" by survey after the enumeration.

Some persons-- like the homeless-- would be missed in an enumeration of households. Special crews (including senior management personnel from the Bureau) counted the homeless in shelters, abandoned buildings, and on the streets, during the night of March 20. Advocates for the homeless have criticized this count as incomplete, but much of the controversy may be definitional: for example, people in half-way houses are counted by the Census in group quarters, not as homeless.

Group quarters, like half-way houses, jails, and college dormitories, are enumerated separately. For example, lists of occupied rooms in dormitories are obtained from colleges; questionnaires are distributed to these rooms and enumerators follow up if there is no response or the number of responses is below the control total supplied by the college.

(3) There are also list/enumerate areas (mainly rural). Unlabeled census forms are distributed by postal carriers; enumerators go through the area, listing addresses, marking them on maps, and collecting forms.

(4) In most rural pre-list areas, enumerators deliver forms, to be mailed back by respondents.

(5) Vacant units are included in denominators of percentages.

(6) Other innovations for 1990: enumerators were allowed to work part time; exemptions were obtained, so persons on AFDC or military pensions could work for the Census without losing benefits.

(7) Another trigger: The Census form has space for seven persons. If the form is completely filled out, there may be eight or more persons in the household, and the respondent did not know what to do; an enumerator calls to check.

(8) MCDs include towns, cities and counties. Tax allocation formulas are quite complicated, and the actual impact of an undercount is a matter of debate; estimates range from a few dollars per person to a few hundred dollars.

(9) Results from "local review" drew headlines. The Bureau gives preliminary counts of housing units to all 39,000 MCDs for review; there were 6,600 challenges, based on local administrative records. However, local records are often out of date, and few of the challenges seem well founded. For example, the City of Boston claimed the Bureau had missed 300 housing units on a block that turned out to be a park--the Boston Common.

(10) See e.g. J. Gleick, *New York Times Magazine*, pp22-27, July 15, 1990.

(11) C.F. Citro, M.L. Cohen, eds., *The Bicentennial Census: New Directions for Methodology in 1990* (National Academy Press, Washington, D.C., 1985); chapter 4 dscusses evaluation and chapter 7, adjustment.

E.P. Ericksen, J.B. Kadane, *J. Amer. Statist. Assoc.* 80, 98 (1985).

D.A. Freedman, W.C. Navidi, *Statistical Science* 1, 1 (1986).

E.P. Ericksen, J.B. Kadane, J.W. Tukey, *J. Amer. Statist. Assoc.* 84, 927 (1989).

R.J. Beran and 12 other statisticians, *Statement on Census Adjustment*, presented to the U.S. House of Representatives Subcommittee on Census and Population, March 3, 1988.

(12) The 1980 case was decided in 1987: *Cuomo et al. v. Baldrige et al. 674 F. Supp. 1089 (S.D.N.Y.)*. Some excerpts from the digest and opinion:

"State, city, and their officials brought action against Secretary of Commerce, Director of the Bureau of the Census, and other officials seeking statistical adjustment of 1980 decennial census. The District Court, Sprizzo, J., held that state and city failed to establish that statistical adjustment of decennial census was technically feasible."

"....it is essential to any such adjustment that a technically feasible adjustment methodology exist which gives a truer picture of the United States population on a state-by-state basis for apportionment purposes, and a sub-state-by-sub-state basis for federal funding purposes....If it does not, then no adjustment can or should be made...because...both congressional seats and revenue sharing funds are fixed quantities, and an increase in the population in one state or sub-state area will adversely affect the shares of other localities....

"Notwithstanding the complexity of the facts.... this action presents one issue to be resolved by the Court: whether the plaintiffs have sustained their burden of proving that a statistical adjustment of the 1980 census will result in a more accurate picture of the proportional distribution of the population of the United States on state-by-state and sub-state-by-sub-state basis than the unadjusted census. The Court finds as a matter of fact that the plaintiffs have not sustained that burden, and the action must therefore be dismissed...."

(13) *City of New York et al. v. Department of Commerce et al., Docket No. 88 Civ. 3474 (U.S. District Court, E.D.N.Y.).*

The 1990 suit was filed before the Census was taken, and in the Eastern District of New York (Brooklyn), rather than the Southern District (Manhattan). In 1980 as well as 1990, New York City was represented by Cravath Swaine and Moore, a leading commercial litigation firm.

Pursuant to the settlement, draft guidelines for making the decision on adjustment were published by the U.S. Department of Commerce (*Federal Register* 54, 51002, Dec. 11, 1989) and comments were solicited. The comments, responses, and final guidelines were also published (*Federal Register* 55, 9838, March 15, 1990). The court upheld the guidelines against a challenge by New York, which sought an order declaring them invalid. The court ruled too that adjustment was in principle constitutional. 739 *F. Supp.* 761 (E.D.N.Y. 1990).

(14) R.E. Fay, J.S. Passel, J.G. Robinson, C.D. Cowan, *The Coverage of the Population in the 1980 Census*. (U.S. Department of Commerce, Government Printing Office, Washington, D.C., 1988). In 1980, demographic analysis series DA-2 estimated the national undercount rate for whites-and-other-races as 0.7 of 1%; the rate for blacks was 5.9%; blacks were about 12% of the population: Table A.80.3.

(15) Estimates of the national undercount rate for non-black hispanics in 1980, made by capture-recapture methods, ranged from 0 to 8%; see (14), Table 7.1.

(16) R. Warren, J.S. Passel, *Demography* 24, 375 (1987); Fay et al. (14).

(17) J.S. Passel (1990). *Demographic Analysis: A report on its utility for adjusting the 1990 Census*. Technical Report, Urban Institute, Washington, D.C. Changes mainly affect children below the age of 10 in 1990.

(18) The adjustment is done by capture-recapture. Capture is in a sample of birth certificates; recapture, in the census. See *Vital Statistics of the United States, 1960 Volume I—Natality*. U.S. Department of Health, Education, and Welfare. Figure 1.1 on p1.1 shows the estimated completeness of birth registration data.

There are more recent tests based on the Current Population Survey. See *Test of Birth Registration Completeness 1964 to 1968*. (U.S. Department of Commerce, 1973).

(19) P.E. Whelpton, *Vital Statistics, Special Reports, Selected Studies,* Vol. 33 No. 8, U.S. National Office of Vital Statistics (1950).

(20) The estimates depend to some extent on the assumed number of illegals in 1980, taken as 3 million. Data are from Fay et al. (14).

(21) Ericksen and Kadane (11).

(22) Passel (17); J.G. Robinson, P. Das Gupta, and B. Ahmed, *A case study in the investigation of errors in estimates of coverage based on demographic analysis: black adults aged 35 to 54 in 1980*. Technical Report, Bureau of the Census. Passel p.27 suggests that black males age 15-19 may be easy to count, because many of them are in institutions (e.g., army barracks). However, about 90% of black males age 15-19 are outside institutions, and the

undercount rate for this non-institutional group is under 1%, which is remarkably low. Passel argues too that demographic analysis can be modified to eliminate the artifact in figure 1, without much change in the estimated undercounts. That may only demonstrate the flexibility of the apparatus.

(23) Passel (17) is more optimistic. Also see J.G. Robinson, P. Das Gupta, and B. Ahmed, *Evaluating the quality of estimates of coverage based on demographic analysis,* Technical Report, Bureau of the Census.

(24) There are many models for capture-recapture; e.g. C.D. Cowan, D. Malec, *J. Amer. Statist. Assoc.* 81, 347 (1986).

(25) Strata are defined on the basis of geography and demographic characteristics from the 1980 Census.

(26) The PES cannot start until Census fieldwork is virtually complete, to preserve independence.

(27) Census data are kept confidential. But many people seem not believe this, and are reluctant to cooperate because they do not want their forms turned over to other government agencies-- like the I.R.S. Asking for social security numbers is likely to make such respondents even less cooperative.

(28) Illustrative citations are given on error rates in matching: errors of several percentage points in size are not unusual; newer algorithms may be better. Reports from TARO and the Dress Rehearsal (38) suggest that even with good data, matching errors will be 1% or more.

T.Z. Hambright, *Comparability of Marital Status, Race, Nativity, and Country of Origin on the Death Certificate and Matching Census Record: United States-May-August 1960*. U.S. Department of Health, Education, and Welfare, N.C.H.S. Series 2 Number 34. (U.S. Government Printing Office, Washington, D.C., 1969)

J.D. Curb, C.E. Ford, S. Pressel, M. Palmer, C. Babcock, C.M. Hawkins, *Amer. J. Epidemiol*. 121, 754 (1985).

C.E. Johnson, Jr., *Consistency of reporting of ethnic origin in the Current Population Survey*. Technical Paper No. 31, Bureau of the Census, U.S. Department of Commerce, 1974.

E. Rogot, M. Feinleib, K.A. Ockay, S.H. Schwartz, R. Bilgard, J.E. Patterson, *Amer. J. Publ. Health* 73, 1265 (1983).

M.A. Jaro, *J. Amer. Statist. Assoc*. 84, 414 (1989).

(29) In 1980, 6 different sets of imputations were considered; and there were two "P-samples" (April and August CPS) for recapture. National undercount estimates ranged from -1.0% to 2.1%: see Table 7.1 in (14). Models have impact. (For present purposes, reweighting is considered an imputation.)

(30) An "erroneous enumeration" is a person counted in error, e.g., a baby born on April 2. Some people are double-counted, e.g., at their regular homes and at vacation homes. A "fabrication" is made up, or "curbstoned" by an enumerator. Failure to take these problems into account could lead to an over-estimate of the population. Similar problems in PES data are not dealt with; such errors, as well as matching errors and bad address information all tend to bias the estimated population upwards. "Correlation bias" (dependence between capture and recapture or heterogeneity in recapture probabilities) may be a partial offset; so far, all attempts to quantify this bias at the local level seem quite speculative.

(31) Census records are searched for a match in the block where the PES person is found, and surrounding blocks. (For movers, the search starts in the block where the person is reported to have lived on Census day.) To avoid bias, levels of effort on the E-side (e.g. to identify duplications or erroneous enumerations) should correspond to those on the P-side, and search areas should be comparable. However, virtually all the potential erroneous enumerations are sent to follow-up; many gross omissions do not go to follow-up.

(32) Movers in the PES are weighted up according to the block in which they are found, but contribute to the estimated population and number of matches according to reported place of residence on April 1, 1990.

(33) The demographic categories are not fully crossed with the geographical stratification: certain cells are collapsed at the end. Five Asian post strata (crossed by age, sex and tenure) were added in fall 1990, following protests by the Asian community. The original design would have made undercount rates for Asians match those for whites-- by assumption. The heterogeneity problem remains for other groups.

(34) It is also proposed to adjust the factors so that at the national level the capture-recapture estimates agree with results from demographic analysis.

(35) For the 1980 Census, it was possible to quantify the impact of assumptions on estimates of undercount and standard errors, at least to some extent. Some observers found that results depended strongly on assumptions; others disagreed (11,12).

(36) Typically, this process leads to fractional numbers of people in blocks, which are rounded to integers before proceeding. Rounding is "controlled" so that totals do not change.

(37) For 1980 data on overcounts, see Fay et al. (14). Their Table D1 shows that in 1980, for white-and-other-race males, PEP 2-9 estimated an overcount in 7 out of 17 age groups (45 and up). The pattern for women was similar (Table D2). From Table 7.5, PEP 2-9 estimated an overcount for 11 out of 50 states. PEP 2-9 was the particular version of the DSE favored by plaintiffs' experts in 1980 (11,29).

(38) G. Diffendahl, *Survey Methodology* 14, 71 (1988); N. Schenker, *Survey Methodology* 14, 87 (1988); H. Hogan, K. Wolter, *Survey Methodology* 14, 99 (1988). C.D. Jones, *Evaluation of TARO*. Technical Report, Bureau of the Census (1988).

Census notes page 13, note 38 continued

The "Dress Rehearsal" in 1988 afforded another test of adjustment methodology; again, the data may be interpreted in different ways.

M.M. Mulry and K. K. West, *Evaluation Follow-Up for The 1988 Post Enumeration Survey*. Technical Report, Bureau of the Census (1990).

M. Mulry and B. Spencer, *Total error in PES estimates of population: The Dress Rehearsal Census of 1988*. Technical Report, Bureau of the Census (1990).

(39) See e.g. Hogan & Wolter (38) pp105-106.

(40) I testified for the Department of Commerce in the 1980 Census adjustment cases, and am working for the Department of Justice on the 1990 case. This article expresses my views, which may not be shared by the Department of Justice. L. Bazel (San Francisco), L. Breiman (Berkeley), P. Diaconis (Harvard), E. Hoag (California Department of Finance), S. Klein (RAND), W. Kruskal (Chicago) and A. Tversky (Stanford) all made useful comments. So did P. Bounpane, G. Robinson, C. Jones, and J. Thompson (Bureau of the Census). J. Passel (Urban Institute) also deserves thanks. Support from the Miller Institute is gratefully ackowledged.

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review. Statistical Science, 1988, Vol. 3 No. 2 239-271.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984, 579-611.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data. Scandinavian J. Statist., 1988, 15, 1-23.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss. Ann. Stat. 16, 483-520, 1988.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data. Biometrika, 1987, 74, 799-808.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition). In Statistical decision theory and related topics IV, Vol. 2, 163-175. Berger and Gupta editors. Springer Verlag. 1988.

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. Statistical Science, 1988, Vol. 3 No. 1, 3-56.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (Nov 1986, revised Aug 1988). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations. T.K. Dijkstra, ed. On Model Uncertainty and Its Statistical Implications. Lecture Notes in Econometric and Mathematical Systems, No. 307 Springer, Berlin pp.1-16.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model. J. Amer. Statist. Assoc., 1988, 83, 744-749.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. Journal of Theoretical Probability, Oct. 1988, Vol. 1 No. 4, 381-410.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery. SIAM J. Appl. Math., June, 1989.

95. CANCELLED

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. Environmental Health Perspectives, 1989, 81, 169-188.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald. Internat. Statist. Review, 58, 129-152, 1990.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (Oct. 1987, revised Aug. 1990). Meeting times for independent Markov chains.

119. HESSE, C.H. (Oct. 1987, revised Nov. 1990). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. Statistics a Guide to the Unknown, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles. Appeared Annals of Statistics, 1416-1437. September 1990.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (Nov. 1987, revised Sept. 1990). Modelling Sedimentation.

126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (Dec. 1987, revised June 1989). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes. Ann. Statist. 18, 1188-1202.

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. CANCELLED

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). Ex-smokers and the multistage model for lung cancer. Epidemiology, Jan. 1990, Vol. 1 No. 1, 21-29

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities. To appear in Ann. Statist.

137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.

138. DONOHO, D.L. and LIU, R.C. (Jan. 1988, revised Jan 1990). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (Dec. 1987, revised Oct. 1990). Statistical aspects of neural networks.

141. BRILLINGER, D.R. (Jan. 1988). Two reports on trend analysis: a) An elementary trend analysis of Rio negro levels at Manaus, 1903-1985. b) Consistent detection of a monotonic trend superposed on a stationary time series.

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density. Ann. Stat., December, 1988.

143. DALANG, R.C. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (Sept. 1989, revised Oct. 1990). On min-max majority and deepest points.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (Mar. 1988, revised Feb. 1990). Asymptotics for multivariate trimming.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). Cauchy's equation and de Finetti's theorem. To appear in Scand. J. Statist.

153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.

154. DOKSUM, K.A. and GASKO, M. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.

155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.

156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.

157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.

158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.

159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.

160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.

161. BREIMAN, L. and PETERS, S. (June 1988, revised March 1990). Comparing automatic smoothers (A public service enterprise).

162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.

163. Cancelled

164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.

165. HESSE, C.H. (July 1988, revised Sept. 1990). On a class of stable process models in economics.

166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.

167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.

168. LE CAM, L. (Aug. 1988, revised Jan. 1990). Maximum likelihood an introduction. Internat. Statist. Review 58, 153-171, 1990.

169. BREIMAN, L. (Aug.1988, revised Dec. 1990). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error.

170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.

171. STONE, C.J. (September 1988). Large-sample inference for logspline models.

172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.

173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.

174. YOR, M. (October 1988). Interwinings of Bessel processes.

175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.

178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subproblems.

179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.

180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. To appear in Proc. Amer. Math Soc.

181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989, Aug. '89). Models for variable stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.

182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.

183. BERAN, R. (November 1988). Calibrating prediction regions.

184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.

185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.

186. Cancelled.

187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.

188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.

189. PITMAN, J.W. and YOR, M. (Nov. 1990). Arcsin laws and interval partitions derived from a stable subordinator.

190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.

191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.

192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.

193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.

194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.

195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.

196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.

197. BREIMAN, L. and SPECTOR, P. (Mar. 1989, revised June 1990). Submodel selection and evaluation in regression — the X-random case.

198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.

199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, Actas del 2° congreso lantinoamericano de probabilidad y estadistica matematica, Caracas, 1985.

200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.

201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.

202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.

203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.

204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p-Balls.

205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.

206. MILASEVIC, P. and NOLAN, D. (May 1989). Estimation on the sphere: A geometric approach.

207. SPEED, T.P. and YU, B. (July 1989, rev. Oct. 1989). Stochastic complexity and model selection: normal regression.

208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.

209. BREIMAN, L. (July 1989, revised Aug. 1990). Fitting additive models to regression data: Diagnostics and alternative views.

210. PEMANTLE, R. (July 1989). Vertex-reinforced random walk.

211. LE CAM, L. (August 1989). On measurability and convergence in distribution.

212. FELDMAN, R.E. (July 1989). Autoregressive processes and first-hit probabilities for randomized random walks.

213. DONOHO, D.L., JOHNSTONE, I.M., HOCH, J.C. and STERN, A.S. (August 1989). Maximum entropy and the nearly black object.

214. DONOHO, D.L. (Aug. 1989, revised Sept., Nov. 1989). Statistical estimation and optimal recovery.

215. STONE, C. (Aug. 1989, revised Dec. 1990). Asymptotics for doubly-flexible logspline response models.

216. NOLAN, D. (August 1989). The excess mass ellipsoid.

217. FREEDMAN, D.A. (August 1989). Statistical models and shoe leather.

218. BICKEL, P.J., NAIR, V.N. and WANG, P.C.C. (August 1989). Nonparametric inference under biased sampling from a finite population.

219. DALANG, R.C. and WALSH, J.B. (September 1989). The sharp Markov property of the brownian sheet and related processes.

220. BURMAN, P. and NOLAN, D. (Oct. 1989). Location-adaptive density estimation and nearest-neighbor distance.

221. LOW, M. (Oct. 1989). Non-existence of an adaptive estimator for the value of an unknown probability density.

222. LOW, M.G. (Oct. 1989). Invariance and rescaling of infinite dimensional Gaussian shift experiments.

223. LOW, M.G. (Oct. 1989). Lower bounds for the integrated risk in nonparametric density and regression estimation.

224 BREIMAN, L. and CUTLER, A. (Oct. 1989). A deterministic algorithm for global optimization.

225. LOW, M.G. (Oct. 1989). Local convergence of nonparametric density estimation problems to Gaussian shift experiments on a Hilbert space.

226. LEHMANN, E.L. (Oct. 1989). Model specification: The views of Fisher and Neyman, and later developments.

227. GOLUBEV, G.K. and NUSSBAUM, M. (Nov. 1989). A risk bound in Sobolev class regression.

228. STARK, P.B. (Dec. 1989). Rigorous computer solutions of infinite-dimensional inverse problems.

229. EVANS, S.N. and PERKINS, E. (December 1989). Measure-valued Markov branching processes conditioned on non-extinction.

230. EVANS, S.N. (December 1989). The entrance space of a measure-valued Markov branching process conditioned on non-extinction.

231. BREIMAN, L. (Dec. 1989, revised July 1990). The Π-method for estimating multivariate functions from noisy data.

232. SHAFFER, J.P. (December 1989). Probability of directional errors with disordinal (qualitative) interaction.

233. BRODSKY, M. and PANAKHOV, E. (January 1990). Concerning a priori estimates of solution of the inverse logarithmic potential problem.

234. SHAFFER, J.P. (January 1990). The Gauss-Markov theorem and random regressors.

235. SPLAWA-NEYMAN, J. (January 1990). On the application of probability theory to agricultural experiments. Essay on principles by Jerzy Splawa-Neyman. (Roczniki Nauk. Bol. Tom. X (1922)). Translation from the Polish original of §9 pp.29-42 by D.M. Dabrowska, edited by T.P. Speed.

236. DONOHO, D.L. and NUSSBAUM, M. (January 1990). Minimax quadratic estimation of a quadratic functional. Appeared Journal of Complexity, 290-323. September 1990.

237. DONOHO, D.L., JOHNSTONE, I.M., STERN, A.S. and HOCH, J.C. (January 1990). Does the maximum entropy method improve sensitivity? Appeared Proc. Nat. Acad. Sci. U.S.A., 5066-5068. July 1990.

238. KOOPERBERG, C. and STONE, C.J. (Feb. 1990, revised Dec. 1990). A study of logspline density estimation.

239. HESSE, C.H. (March 1990). Hitting-time densities of a two-dimensional Markov process.

240. EVANS, S.N. (March 1990). Trapping a measure-valued Markov branching process conditioned on non-extinction.

241. YU, B. and SPEED, T.P. (March 1990). Stochastic complexity and model selection II. Histograms.

242. BICKEL, P.J. and MILLAR, P.W. (March 1990). Uniform convergence of probability measures on classes of functions.

243. DALANG, R.C. and WALSH, J.B. (April 1990). The sharp Markov property of Lévy sheets.

244. BRILLINGER, D.R. (March 1990). Two reports on the analysis of spatially aggregate data:
a) Mapping aggregate birth data
b) Spatial-temporal modelling of spatially aggregate birth data.

245. ZHANG, P. (April 1990). Variable selection in non-parametric regression with continuous covariates.

246. BICKEL, P.J. and ZHANG, P. (April 1990). Variable selection in non-parametric regression with categorical covariates.

247. EVANS, S.N. (April 1990). Association and infinite divisibility for the wishart distribution and its diagonal marginals.

248. FREEDMAN, D.A., KLEIN, S., SACKS, J., EVERETT, C. and SMYTH, C. (April 1990). Ecological regression and voting rights.

249. DONOHO, D.L. and LOW, M.G. (May 1990). Renormalization exponents and optimal pointwise rates of convergence.

250. HESSE, C.H. (May 1990). The one-sided barrier problem for an integrated Ornstein-Uhlenbeck process.

251. HESSE, C.H. (May 1990). Rates of convergence for the empirical distribution function and the empirical characteristic function of a broad class of linear processes. To appear - J. Multivariate Anal. (Nov. 1990).

252. EVANS, S.N. (May 1990). Equivalence and perpendicularity of local field Gaussian measures.

253. SPEED, T.P. (May 1990). Introduction to "The Arrangement of Field Experiments" by R.A. Fisher (J. Min. Agric. Gr. Br., 33, 503-513; Collected Papers II, 38: 83-94).

254. BRODSKY, M. (May 1990). Physically meaningful results on the uniqueness of inversion of Newtonian potential.

255. BICKEL, P.J. and RITOV, Y. (May 1990). Estimating linear functionals of a PET image.

256. BICKEL, P.J., RITOV, Y. and WELLNER, J.A. (May 1990). Efficient estimation of linear functionals of a probability measure P with known marginal distributions.

257. ZHANG, P. (June 1990). Model selection via multi-fold cross validation.

258. BICKEL, P.J. and FAN, J. (June 1990). Some problems on the estimation of densities under shape restrictions.

259. DONOHO, D.L. and LOW, M.G. (June 1990). White noise approximation for sampled data.

260. YU, B. (June 1990). Rates of convergence and central limit theorems for empirical processes of stationary mixing sequences.

261. STARK, P.B. (July 1990). Inference in infinite dimensional inverse problems: Discretization and duality.

262. EVANS, S.N. (August 1990). Polar and non-polar sets for a tree indexed process.

263. LE CAM, L. (August 1990). Some recent results in the asymptotic theory of statistical estimation.

264. BARLOW, M.T., EVANS, S.N., and PERKINS, E.A. (August 1990). Collision local times and measure-valued processes.

265. LE CAM, L. (August 1990). Some special results of measure theory.

266. BASAK, G.K. (August 1990). A class of limit theorems for singular diffusions.

267. BASAK, G.K. and BHATTACHARYA, R.N. (August 1990). Stability in distribution for a class of singular diffusions.

268. STONE, C.J. (August 1990). L_2 rate of convergence for interaction spline regression.

269. LE CAM, L. (September 1990). An infinite dimensional convolution theorem.

270. BICKEL, P.J., RITOV, Y. and WELLNER, J.A. (Sept. 1990). Efficient estimation of linear functionals of a probability measure P with known marginal distributions.

271. DONOHO, D.L. and LOGAN, B.F. (September 1990). Signal recovery and the large sieve.

272. BICKEL, P.J. (October 1990). Some theory for the stringer bound of auditing practice.

273. BICKEL, P.J. (October 1990). Theoretical comparison of bootstrap t confidence bounds.

274. HESSE, C.H. (October 1990). Further refinements of a stochastic model for particle sedimentation in fluids.

275. LOW, M.G. (October 1990). Renormalizing upper and lower bounds for integrated risk in the white noise model.

276. KLEIN, S., SACKS, J. and FREEDMAN, D.A. (October 1990). Ecological regression versus the secret ballot.

277. GUERRA, R. and SPEED, T.P. (November 1990). Statistical methods for analyzing DNA-DNA hybridization data.

278. LEHMANN, E.L. and ROJO, J. (November 1990). Invariant location orderings.

279. FREEDMAN, D.A. (December 1990). The 1990 census and statistical adjustment.

280. BJERVE, S. and DOKSUM, K. (December 1990). Correlation curves: Measures of association as functions of covariate values.

281. DONOHO, D.L. and JOHNSTONE, I.M. (December 1990). Wavelets and optimal nonlinear function estimates.

282. DONOHO, D.L. (December 1990). Gel'fand n-widths and the method of least squares.

283. FAN, J. and BICKEL, P.J. (December 1990). Nonparametric density estimation: A piecewise maximum likelihood approach.

284. FREEDMAN, D.A (January 1991). Nonparametric binary regression: a Bayesian approach.

285. DONOHO, D.L. (January 1991). Super-resolution via sparsity constraints.

286. BRILLINGER, D.R. (January 1991). Some asymptotics of finite fourier transforms of a stationary p-adic process.

287. BRILLINGER, D.R. and KAISER, R. (January 1991). Fourier and likelihood analysis in NMR spectroscopy.

288. SPEED, T.P. (January 1991). Contribution to the discussion of: "Interpreting blocks and random factors" by Myra L. Samuels, George Casella and George P. McCabe.

289. BRODSKY, M. and STZAKHOV, V.N. (January 1991). On the uniqueness of determination of a polyhedron by its edges with applications to the inverse potential problem.

290. SPEED, T.P. (January 1991). Notes towards a critique of methods of estimating bias and variance in undercount estimates.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.