Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TECHNICAL EXPERT WORKING SESSION
PROCEEDINGS

IMPROVING SES ESTIMATORS

TABLE OF CONTENTS

IMPROVING SES ESTIMATORS

EXECUTIVE SUMMARY

Education research has relied for over half a century on eligibility for free/reduced price lunch (FRL) as a primary indicator of a student's socioeconomic status (SES). With changes in the regulation and the implementation of FRL practices, it is no longer a stable indicator with a universal and relevant definition.

The working session engaged technical experts and NCES staff in discussion of criteria for development and evaluation of new indicators to replace FRL in research and reporting. Attention then turned to consideration of a proposal to use geographically-based information linked to available survey data from the American Community Survey.

The experts and the NCES staff discussed options and potential consequences of various approaches with regard to:

- the purposes for which the new index would likely be used,
- a geographic basis and the availability of demographic information for geographic units,
- basic units of application - student, school, neighborhoods, geographic units,
- data elements for calculation of index values - based on household composition and information,
- data elements for calculation of index values - income or economic information,
- potential problems and pitfalls, both pragmatic and technical,
- imputation of neighborhood similarity,
- statistical methodology including kriging, and
- requirements for appropriate vetting of a new index proposal.

A final list of 50 key points related to these nine areas of discussion led to the following recommendations from the assembled experts.

INDEX DEFINITIONS

Geographic Basis: Use the GIS system in formulating the indices and prepare output in map, table and text forms.

Purely Economic Index: Base the index on income rather than wealth; ensure Title 1 specifications (including adjusting total family income by family size).

Broad Socio-Economic Index: Compose the index from three elements: 1) Income - Total household income, adjusted by number of occupants, 2) Education level - maximum for adults in household, and 3) Social status - highest occupational status index value. These data are available from the ACS survey.

Precision: Precision of each index must be transparent - first as an index, especially as used for aggregates (e.g., distribution of index values for a school); and second, with respect to ascribing values to individual addresses.

Vetting the Index: Invest in the design of exhaustive vetting, with early stage decision-making supported by experimentation, simulation and testing.

Technical Expertise: Acquire or engage advanced technical and statistical expertise, whether as a panel of experts or as individuals, to ensure the technical validity and credibility of the indices being developed.

IMMEDIATE NEXT STEPS

Conduct research on open issues.

Investigate related indices for inclusion in some part of the development process or for comparison.

Develop a comprehensive statistical strategy.

Design vetting in order to be alert in planning to identify potential omissions from index construction or other weaknesses and critical properties (to be built into the index formulation).

PREFACE

The National Center for Education Statistics (NCES) charged the National Institute of Statistical Sciences (NISS) with convening a diverse set group of experts on creation and utilization of socioeconomic status (SES) indices to provide background on the technical, social and economic requirements. On 4-5 March 2019 the technical experts met in person with NCES staff to discuss a new proposal for a new SES index to be useful for education researchers and education policy-makers.

IMPROVING SES ESTIMATORS

I. INDEX PURPOSE

Objective

To create a new measure of socioeconomic status that is suitable for use with education data.

Charge to Workshop

This workshop provides a forum to discuss the critical issues in creating a new geographically-based index of socioeconomic status utilizing federally available data – not just a poverty indicator – that is sufficiently robust to be stable and reliable as populations shift over time. The purpose of the workshop is to evaluate progress to date in development of a new index and to consider next steps.

Issues discussed in detail should include the known and likely uses for the new index, components of the new index, methodology for index calculation, assessment of potential vulnerabilities and requirements for vetting the index.

II. INDEX ELEMENTS

Starting Point – Strengths and Weaknesses of FRL

For decades, eligibility for free/reduced price lunch (FRL) was the surrogate economic indicator used for most education research and also was used for other research in economics, sociology and other social sciences. While changes in the definition of "eligibility" have now rendered universal use of FRL eligibility unworkable, it is worth examining FRL as originally conceived in order to identify particular attributes to preserve in the new index.

Uses of FRL

Administration:

- Information source for making education policy.
- Resource for determining allocation of resources.
- Metric meeting federal and state reporting requirements.

Research:

Research using education data.

- Cross-sectional and longitudinal research on schools and on individuals.
- Evaluation of interventions (for disadvantaged students).

FRL Definition

Index scale:

- Poverty metric - usually binary (FRL eligible or not) but capable of distinguishing between two low-income groups: Free or Reduced-price eligibility.

Metric definition:

- US (USDA) poverty definition (scaled for families, by size) Note: 1939 definition of "family."
- Nation-wide definition with no adjustment for local economics or for self-sustaining capability (e.g., farming/ranching).

Index unit:

- Individual student.

FRL Information and Accuracy

Inclusion:

- Self-designation (with documented bias against inclusion by post-elementary students).
- Designation attached to student's record.

Timeliness:

- Updated annually.

Information Accessibility:

- Student-level record maintained at student's school, also by district (sometimes as part of state records).
- School-level summary information (e.g., %FRL students) – public record.

INDEX ELEMENTS – *SIDE* PROPOSAL

Uses of Index

Research:

- Research using education data, cross-sectional studies, longitudinal studies of schools but not of students.

Administrative Decisions:

- School-level information as source for making education policy.
- Resource for determining allocation of resources.
- Student and school level information meeting federal and state reporting requirements.

Index Definition

Index scale:

- 10-point scale based on US national deciles for household income, i.e., equal numbers of US households per classification.

Metric definition:

- Nation-wide definition with no adjustment for local economics or for self-sustaining capability (e.g., farming).
- Distribution of student population by income deciles.

Unit of application:

- Finer than school or district.
- Individual student's immediate geographic neighborhood, not student's address-based.

Index Calculation

Data resource:

- American Community Survey (ACS).
- Potential for incorporating available data from other federal data bases.

Geo-neighborhood:

- GIS grid with 300m x 300m squares.
- Geographically comprehensive covering 50 states, Puerto Rico, Guam.

Methodology:

- Kriging - proposed, subject to evaluation of suitability.

Index Information and Accuracy

Inclusion:

- Data sources (not self-designation) - geographically comprehensive.
- Estimates for areas with resident students.
- Imputation for areas currently without student residents.

Timeliness:

- Annually based on American Community Survey 5-year rolling sample.
- Information accessibility:
- Student-level record maintained at student's school, also by district (sometimes as part of state records).
- School-level summary information (e.g., frequency (or %) distribution by decile) – public record.
- Fine-scale map for neighborhoods as well as numeric data.

Accuracy and Precision:

- Student-level: Within a 3-decile span (i.e. assigned decile +/- 1 decile).

III. ISSUES FOR DISCUSSION

Premise

Replacing FRL requires a new index at the individual student level that is universally applicable on a national scale; however, it will not be feasible to obtain the needed socioeconomic information directly. Rather, the index will need to be constructed using administrative information that is consistently defined and recorded throughout the US. While the FRL functioned as a (binary) poverty index, its replacement should function as a full-spectrum economic index.

A new index must be nationally applicable, allowing annual updating of student/school records. Its formulation must be transparent and without discernible bias; it must be computationally feasible and cost effective.

The American Community Survey (ACS) is a source of comprehensive socio-and-economic data at the family/household level; it is conducted nation-wide and refreshed annually so that information remains current. Further, ACS records are geo-located allowing GIS mapping at the unit level. Students' residential addresses, required as part of student record information, provide the link via GIS to (federal) geo-located administrative data including unit level ACS data.

Using a GIS framework offers a huge potential advantage to constructing a new index. This means that in addition to ACS data, data from many other sources could also be incorporated into the process of defining the index and into analyses of education research questions.

Purpose(s)

A new index is needed to serve several distinct purposes at different levels of aggregation.

Key Points:

1. States and school districts require a measure that meets federal reporting requirements and that will be used in allocating both federal and state funding. The index may also be used in classify schools (e.g., NAEP reports).

   Policy makers (and researchers) are interested in evaluating policies and programs conducted at the school level – for example analysis of test scores following increased spending to target a specific need or disadvantaged group.

   Researchers and policy makers studying outcomes or changes over time or regional differences for students with differing backgrounds may be able to analyze results for groups of students. However, this will not extend to analyses of individual student data because the index applies only to the neighborhood of the student's residence, not necessarily to the student.

2. Schools, districts and states are in need of a measure and will use what is available to them – even if it is not the best measure. Hence uses and potential for misuses of the index (e.g., gaming the system) must be thoroughly considered and evaluated.

3. There is a difference between SES and poverty. Some identified purposes require a narrower economic index (EI), for example to meet certain federal reporting requirements. High priority uses of poverty measures at the school level include funding and resource allocations. Uses of poverty measures at the individual level would require additional individual student information from other sources.

4. Other identified research and policy purposes, for example analysis of student data and prediction of outcomes, depend on a broader measure of student disadvantage rather than simply poverty.

Recommendation:

Construct two indices: 1) a narrow purely Economic Index (EI) specifically constructed to meet federal (and state) reporting and classification requirements; and 2) a broader Socio-Economic Index (SEI) that encompasses social as well as economic factors and uses contemporary definitions and measures of factors of both.

Index Development Strategy

The index development has five aspects: GIS basis and the geographical unit a raster cell identifies; index content, i.e., components or ACS items, and index scale; definition of raster cell neighbors (for purposes of model-based estimation); statistical model with estimators for uncertainty; evaluation for accuracy, reliability, stability and vetting for multiple uses.

Key Points:

5. The GIS basis has already been decided to be 300 x 300 m cells on a raster that covers the US – 50 states, the District of Columbia, Puerto Rico and Guam. There are many non-congruent geographic partitionings in use that are relevant to education data: school district, individual school boundaries, census blocks or tracts, state/county/town or township political boundaries and defined economic areas, etc. Since each cell is smaller than any of these partitionings, it can simultaneously and independently be assigned values from multiple differently partitioned data files. For each data file, the cell takes the value based on the particular partition it belongs to, creating a vector of values for the elements in the multiple files.
6. Each proposed index is intended to be used as a full-spectrum index, not simply a binary indicator (e.g., poverty index). Without individual reporting, the levels of precision and accuracy (on an individual student scale) will not be as high as with personal reports. Consequently, the accuracy of a categorical index is limited this case - in this case restriction to 10 categories is proposed (other number of categories could be tested and compared).

Unit(s) of Application

The unit of application is both the unit for which the index is calculated and to which it is associated. The index may also be calculated for larger units or aggregations. In this case the unit is the student's neighborhood and the index therefore is not necessarily accurate when applied to the individual student.

Key Points:

7. By defining a new index based on students' residential addresses, the index is a measure of the neighborhood socio-economic level or of "neighborhood disadvantage/poverty."
8. As proposed, both new indices, while calculated for individual students according to their residential addresses, are actually area measures associated with the student's neighborhood. Clarity on this point is important. Otherwise without proper explanation users may make the mistake of assuming these measures can be used for student level estimates.
9. Employing a raster with very fine divisions is proposed to locate each individual student's address in a small neighborhood to minimize within-neighborhood heterogeneity. Zip code, for example, is too coarse to expect reasonable socio-economic homogeneity. Similarly, school district is similarly too expansive.
10. Since many students attend schools outside their immediate neighborhoods or outside their school districts, starting from the geo-location of the school itself is not logical.

Data Elements of the Index – Household/Family

US households with children take many forms in addition to a traditional parents-plus-children family. Definitions relying on the (1939) traditional family structure require reconsideration.

Key Points:

11. The Census Bureau uses a traditional definition of family that does not necessarily match household composition, nor does it include unmarried partners or foster children. Therefore, the proposal to use an income-to-poverty ratio (ACS data) uses the Census Bureau poverty level definition that adjusts for (traditional) family size

12. For the EI, with the purpose of satisfying definitions for federal programs (e.g., Title 1), it makes sense for families to be defined as closely as possible to the way they are defined under the FRL program. Note, however, that with FRL, the individual filling out the paperwork was the one to define the family within the household, while an algorithm will be used with ACS data.

    Departure from the Census definition of family will result in an index (EI) that will not conform to the Census poverty measure.

13. For the SEI, which is not constrained by specific reporting requirements, it is possible to reflect the actual household composition. This is more consistent with research aims and with analyses required for contemporary and relevant policy decisions.

14. Based on different definitions of family/household, EI and SEI will consequently also have different referent individuals when defining social metrics. For EI this will typically be a parent while for SEI it will an adult member of the household will be designated. It will be important for measures should be analyzed to compare those based on household with those using Census measure of family. (See Vetting, below.):
    a) Education level: Highest educational attainment for EI - Parent / SEI - Adult in household.
    b) Occupational Status level: Highest occupation level/social status* for EI - Parent / SEI - Adult in household.
    - *standard listing available - NORC annually updated listing.

15. Education level of adults responsible for children is a crucial metric.
    a) Highest educational attainment for: EI - Parent / SEI - Adult in household.

16. Occupational status is another crucial metric.
    a) Highest occupation status level for: EI - Parent / SEI - Adult in household.

17. Measures (above) should be analyzed to compare those based on household with those using Census measure of family. (See Vetting, below.)

18. For the broader SEI, the goal is not to measure poverty. Rather, the goal is to capture the factors that determine a child's ability to 'succeed' based on their family background. Discussion ranged around the inclusion of these factors in the SEI v research on the impact of these factors beyond an SEI that excluded them. Discussion included such measures as the following:
    a) Number of unemployed parents and/or unemployed adults in the household.
    b) Internet access within the home.
    c) Grandparents raising children.
    d) Seriously disabled family member in the household.
    e) Number of people in the household and number of bedrooms.
    f) Whether a household member receives federal support (such as food stamps).
    g) Presence of running water within the household.
    h) Number of books within the home.
    i) Children's access to museums, summer camps, and music.

19. When considering variables/factors measured independent of poverty (or economic category), it is important to recognize that factor effects cannot be analyzed independently for factors included in the index (e.g., single parent families).

Recommendations:

For the EI, only education level is relevant: Educational attainment of most educated parent.

For SEI, both education and occupational status are relevant: Educational attainment of most educated adult in household and Occupation status of highest status adult in household.

Other social factors should be excluded from SEI to allow analysis of their impacts beyond or independently of the SEI.

Data Elements of the Index – Income

The primary economic indicator for each index is income rather than wealth. The FRL used income rather than wealth, and there appears to be no reason to change this. Income information is also available on an annual basis and therefore can easily be kept current. Of the possible income measures, several were considered: household or family total income, total income adjusted for the number of people included, highest income for an individual. Also considered were other relevant indicators in particular measures of housing costs and local cost of living metrics.

Key Points:

20. In thinking about an essentially – or purely - economic index as a replacement for FRL, comparison could be made on two aspects. FRL served two different roles: a measure of poverty, and a predictor in education research. It is quite possible the new index might outperform FRL in one role but not the other. Testing will be needed to investigate this question.

21. In terms of income, there was a discussion about how to adjust total income, i.e., whether based on the number of adults or the total number of people within the family (for EI) or household (for SEI). If testing of both options shows little difference, the simpler option should be used.

    The more difficult question relates to imputing an income - from an ACS sampled unit or units - to the individual student addresses. The choices are a single ("closest") neighbor, an average value for combined ACS neighbors, a random selection from among ACS neighbors. (Statistical implications of choices are discussed in a later section.

22. For the EI, a categorical measure of income, by covering the income spectrum for the US would provide more information than a binary poverty measure. The proposal to categorize by deciles based on national distribution of incomes would ideally reflect only those incomes for families/households with children, although this might be an unimportant refinement. The precision required would be +/- 1 decile.

    It should be noted that the definition of the income component will have to concord with the information used to create the deciles. Also, the deciles will not necessarily be consistent with other (federal or state) poverty definitions.

23. For the SEI, there are more options as an economic indicator could enter the model directly as one component of the index, or it could enter the model as a covariate (e.g., local cost of living adjustment) or it could be used solely as a covariate in determining/quantifying the similarity of raster cells used to impute for non-ACS-sampled student addresses.

    In any case, it would be desirable to have a continuous primary income measure, although this may pose statistical questions of feasibility. A continuous measure would allow for a finer analysis of how income predicts student outcomes. Precision required (to justify a continuous rather that 10-category measure) would be substantially less than +/- 1 decile (perhaps +/-0.5 decile?).

24. The SEI should include those variables in the narrower EI (i.e., income and education) and build upon it. Discussion focused on measures that could be aggregated to higher levels than the household and could help to explain the variance present in prior models (i.e., those using FRL).
    a) A measure of housing should be included in this index. Possibilities include:
        i. Estimated housing costs: monthly mortgage, housing cost per square foot, ratio of income to mortgage, and/or rental amounts.
        ii. Proportion of renters to owners.
    b) Cost of living indicator: price parity (include rent as well as food), consumer price index (CPI) (can be used to adjust US poverty line).

Recommendation:

An income measure should be a component in both indices. Other indicators, housing cost or local cost of living should be considered for the SEI and certainly retained as covariates for determining or quantifying similarly of raster cells and for use in data analyses.

Caveats and Potential Pitfalls

Caveats for a new index fall into categories: misunderstanding the index itself leading to misuse, and misattribution as for example to individual students rather than students' neighborhoods.

Key Points:

25. Both the new indices need to be transparent: transparent with respect to the unit of application, transparent with respect to precision of each index as used both to a primary unit (student) and to an aggregate (school). An index (EI) built to explore the economics of an area should not be considered an indicator that is predictive of student achievement. Rather, it is a measure of poverty/degree of wealth.
26. One additional reason for discarding the idea of the school location as the reference point is the potential for 'gaming the system" by siting school facilities in order to take advantage of the economic index.
27. Particularly since more than one index is recommended, it is important to note that there is a concern with multiple indices. Users, especially those that are political, will pick the one that best serves their purposes. Therefore, the concordance of the two indices needs to be analyzed carefully and understood. However, this does not mean that they should be identical or entirely consistent at either student or school level. To the contrary, they should be loosely related (they share components) but differ according to their different content and different purposes.

Recommendation:

Careful thought should be given to naming the new indices to indicate as clearly as possible what each index measures.

Imputation and Neighborhood Similarity

In 2017, the five-year accumulated samples for the ACS numbered about 3.54 million residential addresses/housing units. The samples were spread across 3142 counties and county-equivalents in the 50 states and the District of Columbia plus households in Puerto Rico and Guam. With the fineness of the raster, only a very small proportion of the very small raster cells will include ACS sampled households; also, because the raster cell size is fixed, the population will vary greatly among raster cells.

Key Points:

28. This means that index values for the majority of students will be imputed based on raster cells that do not include the students' own residences. However, in very densely populated (urban) areas there may be multiple ACS sampled households within a single raster cell. These circumstances present different questions requiring statistically sound solutions.
29. Imputation for a raster cell with no ACS sampled household poses requires two decisions: how to determine similarity and how many similar cells to include.

    In the current proposal, similarity is defined geographically. Individual student level estimates are formed using the 25 geographically nearest neighbor raster cells with ACS respondent households. The question of including ACS households without children was discussed in detail. Estimates or imputed index values are needed for all raster cells that do or potentially could contain residential addresses. Restriction to ACS respondent households with children limits the number of households from which information can be drawn. It also excludes information for neighborhoods that happen to currently, but not permanently, have no school age children. The number of neighbor cells (25) had been determined empirically, based on a careful test including both larger and smaller numbers.

    Despite Tobler's first law of geography ("Everything is related to everything else, but near things are more related than distant things," (Tobler 1970)), attributes other than geography can contribute heavily to similarity of neighborhoods. For example, urban "gentrification" leads to high-value residential properties adjoining urban decay. Or, two rural areas might be more alike than the 25 geographically nearest neighbors of a single rural area bordering a city especially in the sparsely populated rural areas of the western US. Before adopting a purely geographic approach for assigning index values, differences between a more comprehensive definition of similarity should be evaluated and the consequent differences in ascribed index values should be tested and analyzed.

    Other attributes for use in defining similarity include economic variables that are neighborhood rather than personal economic measures. These could include variables (with federally available data) such as:

    a) household size/number of bedrooms,
    b) % neighbors receiving public assistance,
    c) Population density,
    d) Rent-to-own ratio,
    e) Median contract rent/median house price,
    f) Urban/…/rural,
    g) Median property size, and/or
    h) Price parities by area or localized cost of living (federal data).
30. Educational "goods" include much more than income. For these to be used in education research in conjunction with either new index, they should be excluded from the definition of similarity and excluded as components of the index. While similarity might be construed to be a micro-neighborhood measure of "need" (in the context of education processes or outcomes), elements like motivation, homelife, access to education-relevant resources, should be excluded in defining neighborhood similarity although they may be highly useful as factors or covariates in addressing research questions.
31. For raster cells that include one or more ACS sampled households the decisions are whether, when and how to "borrow" from similar raster cells. (Actual estimation is left for discussion on Statistical Methodology below.)

Recommendation:

Before adopting a purely geographic approach for assigning index values, differences between a more comprehensive definition of similarity should be evaluated. Differences should be tested and analyzed at the individual level and at the aggregate (e.g., school) level.

Statistical Methodology

The discussion revolved around the form of the index to be associated with an individual student, the methodology for combining data from neighboring cells, specifically considering kriging, and finally with representation of variation.

Key Points:

32. For the simpler index (EI), the imputed value for a student's neighborhood could be calculated from a point estimate (e.g., weighted average) of the responses from the similar raster cells with respondents. The result would be assignment to one of 10 deciles.

    Alternatively, instead of assignment to a single category, a set of probabilities (totaling 1) could be distributed across the deciles. At the school level, this would yield a combined distribution of frequencies for the deciles - different in construction but not different in form from aggregating single categories for individual students. The usefulness of this alternative might depend on the ways in which the index is to be used with individual student data.

33. Depending on how the actual estimator or index value is calculated, there can be a problem of loss of randomness. For example, if the collection of similar raster cells is the same for several students, identifying the same index value (best estimate) for all those students neighborhoods removes the variation that is naturally present for any single raster cell if all students in that cell were enumerated and responded directly. As a result, the variation calculated for the index does not accurately represent the variation that is actually present in the population of students represented. This caveat could be addressed by inclusion of a within-raster variation component or by introduction of exogenous randomness into the values ascribed to individual students or by other accepted statistical methods.

    This problem of elimination of variation is even more apparent for a continuous valued measure. It also has implications for calculating variances and standard errors.

34. Kriging was developed for modeling continuous spatial processes. So the application of kriging methods to data that does not arise from a continuous underlying process (continuous spatial support) is a fundamental disconnect.

    It is an open question whether Euclidean distance is a proper or a likely successful framework here. At the same time, it was noted that there are statistical approaches to combining demographic and geographic distances that might be applicable.

    Although Empirical Bayes Kriging touts the ability to encompass spatial change-of-support (correlations of raster cells across a fixed distance differs geographically), this methodology has not been vetted in the peer-reviewed statistics literature. Further, Empirical Bayes Kriging does not lend itself to combining demographic and geographic distances. However, there are thoroughly vetted advanced spatial prediction approaches that do accommodate high-dimensional, non-stationary spatial processes that are suited to a broader range of spatial and data supports. These are likely to be more appropriate and statistically sound.

35. Covariates are needed for any of the approaches mentioned here.

36. There are different roles that covariates can play. Covariates can be incorporated directly into the spatial model, covariates can be used to define relative similarity, covariates can be used in definition of weights. How best to use covariates for the selected modeling approach is an open question that requires thoughtful expert consideration.
37. Another open question to solve is the formulation of variance estimates/standard errors when the sources of variation are quite different. Sampling variation arises from the sampling process – the inclusion/exclusion of individual units from the sample, coupled with probabilities of non-response. Model variation measures deviation from fit. Combining these into a meaningful measure of uncertainty is a conceptual and theoretical statistical problem that will require solution. Similarly, bias can arise from sampling or bias can arise from the model, especially if data are transformed. The problem of finding a useful definition of bias is analogous to the problem of defining uncertainty in this setting.

Recommendation:

Assemble a technical panel or otherwise engage statistical expertise to address key technical issues: the form of the index, the statistical approach and model, roles for covariates, and the definitions of uncertainty and of bias when disparate sources contribute different kinds of variation and bias.

Vetting the Index/Indices

Vetting is one of the most critical issues in the establishment of a new index. The purpose of vetting is to examine in global overview and in micro detail how the index performs. Done well, vetting serves to: 1) identify limitations in applicability or in performance attributes, 2) make choices among components and definitions and to measure the impacts of these choices, 3) assess the index's "fitness for purpose" for intended or likely uses, 4) determine how predictive the index is and to understand what it helps to predict, and 5) determine the index's statistical properties.

Sufficiency of vetting depends on how well the sources of error and of variation are captured by the testing process, how broadly the kinds of applications and uses are covered, and how sensitive the testing approach is.

For example, for this GIS-based design, reproducibility is not likely to be a good measure of accuracy or precision (equivalently, uncertainty). That is because shifting the raster can maximally change the centroid of a single cell by less than ¾ the cell width (300m); and this will not produce much variation (often none) in the index value for that cell. For the same reason, the true variation will not be directly measurable because nearby students will have identical index values assigned, yielding variation = 0, assuming a single assigned index value for the raster cell.

Thus, a comprehensive vetting itself needs to be a carefully designed collection of experiments, analyses and simulations.

Key Points:

38. <u>Intent</u>: to assess the value of the index for use as a full-spectrum economic or socio-economic indicator both for policy-making and for research - at individual student and at the school (or other aggregate) level.
39. <u>Purpose</u>: to identify limitations/to make choices and measure their impacts/to assess "fitness for purpose"/to determine properties of the index.
40. <u>Basic Requirements</u>: feasibility of calculation with updating/refreshing annually on scale of every school district, transparency and reproducibility by analysts, satisfying Title 1 specifications, applicability in both policy and research contexts (at least to the extent that FRL has been used)

41. Resources: simulation of data, historical information, contemporary information, alternative metrics.
42. Levels: student / school / district / other aggregate or subpopulation as defined by demographics or individual or school attribute
43. Outcomes to Examine: known uses for reporting aggregate information, research on individuals, research on predictors of outcomes, high priority uses for poverty measures.
44. Sources for Verification: poverty measures in use in other contexts - see Koedel and Parsons discussion on direct certification, FRL (for elementary students), records of actual incomes, other sources of information outside of ACS.
45. Direct Comparisons: poverty distributions across areas, established relationships e.g., (income to performance or parent education to student performance or poverty distribution for rural v urban),

    Note: In using test scores as an educational performance measure, different tests or scores can be chosen for different grade levels. It should also be noted that optional tests (e.g., SAT or ACT), often present a nonignorable selection bias based on either access or student choice.
46. Comparisons of Predictive Value: participation in federal assistance programs, test scores as measure of education performance (e.g., FRL is predictive of test scores).
47. Aggregates to Vet: commonly used demographic distributions (age, grade % retention by grade) summary statistics (absenteeism, graduation rate, 9th grade GPA), within district socioeconomic indicator index distributions.
48. Choices of Components or Definitions: income v wealth, family v household, Euclidean distance v multi-measure distance metric, alternative methodologies.
49. Metrics for Comparison by Component and by Index Value: agreement (point estimates), precision (uncertainty and magnitude e.g., differences relative to uncertainty), systematic differences by area, region or covariate, variable predictive power within the index.
50. Other Vulnerabilities: misunderstanding of student-level information leading to misuse, potential for use as a gameable index (e.g., for siting an independent school or for deciding student admissions).

Recommendations:

Thought and resources should be invested in the design of a careful collection of experiments and simulations to constitute an exhaustive vetting of proposed indices. Different sizes and different kinds of experiments and simulations will be appropriate for different vetting tasks. These may variously take place over the course of index development and testing; and the initial plan may be sequentially adapted based on earlier vetting results.

Individual experiments within the plan should be designed for efficiency. For example, simulation is suited to evaluation of multiple scenarios, where “truth” is known. On the other hand, small verification experiments may be sufficient when using difficult-to-access data, for example individual household income records or when wanting a rapid screen to reduce the number of alternatives.

A general principle for making choices is to prefer the simpler solution (technically simpler or operationally simpler or simpler to interpret or even more familiar to the user community) when there is only small loss compared to a more complicated or more unusual alternative.

Vetting should be attentive to selective phenomena so that excellent performance of an index at the national or even state level does not mask poor performance on small definable subgroups. (Poor performance for urban enclaves, for example, might affect larger aggregates minimally but would profoundly affect the small numbers of schools involved.)

IV. RECOMMENDATIONS FOR NEW INDICES

New Index Definitions

Geographic Basis: Use the GIS system in formulating the indices and prepare output in map, table and text forms.

Purely Economic Index: Base the index on income rather than wealth; ensure Title 1 specifications (including adjusting total family income by family size).

Broad Socio-Economic Index: Compose the index from three elements: 1) Income – Total household income, adjusted by number of occupants, 2) Education level – maximum for adults in household, and 3) Social status – highest occupational status index value. These data are available from the ACS survey.

Precision: Precision of each index must be transparent – first as an index, especially as used for aggregates (e.g., distribution of index values for a school); and second, with respect to ascribing values to individual addresses.

Vetting the Index: Invest in the design of exhaustive vetting, with early stage decision-making supported by experimentation, simulation and testing.

Technical Expertise: Acquire or engage advanced technical and statistical expertise, whether as a panel of experts or as individuals, to ensure the technical validity and credibility of the indices being developed.

Immediate Next Steps

Research open issues: Investigate related indices for inclusion in some part of the development process or for comparison.

Develop statistical strategy.

Design vetting in order to be alert in planning to identify potential omissions from index construction or other weaknesses and critical properties (to be built into the index formulation).

APPENDICES

Appendix A: Agenda

Appendix B: Background Information

- Abbreviated Background Notes on Kriging
- Abbreviated Background Notes on Small Area Estimation

Appendix C: List of Participants

Appendix D: Biosketches of Technical Experts

Appendix A: Agenda

IES/NCES / NISS Working Session on Improving SES Estimators

March 4-5, 2019 ● Washington, DC

MONDAY, MARCH 4, 2019

9:00am	INTRODUCTIONS
9:30am	SETTING THE STAGE AND DEFINING REQUIREMENTS
11:30am	ROLE OF KRIGING
1:00pm	LUNCH
1:45pm	INPUTS TO THE MODEL (1)
3:15pm	INPUTS TO THE MODEL (2)
4:30pm	WRAP UP
5:00pm	ADJOURN

TUESDAY, MARCH 5, 2019

9:00am	IDENTIFYING OPEN QUESTIONS
11:00am	VETTING THE MODEL
12:30pm	LUNCH
1:30pm	HIGHPOINTS AND HOW TO MOVE FORWARD
4:00pm	ADJOURN

Appendix B: Background Information

Abbreviated Background Notes on Kriging

Kriging is a spatial prediction methodology due to Matheron (1962), with roots in the geodesy work of Gauss carried out early in the nineteenth century. As part of his broader development of the field of geostatistics, Matheron coined the term kriging in honor of D. G. Krige, a South African mining engineer [see Cressie (1990) for an account of the origins of kriging]. This method is well known to geologists, hydrologists, soil scientists, ecologists, and so forth, as a statistical methodology that generates optimal predictions (or interpolations) and associated prediction uncertainties based on the theory of stochastic processes.

The original problem that led to the development of kriging was to interpolate among observations at spatially defined locations in order to create a complete surface representation of the patterns of mineral resource deposits. These predictions or interpolations are optimal in the sense that they are the best linear (weighted) combination of the observations in terms of minimizing the overall mean square prediction error. Importantly, the weights applied to each observation are proportional to the covariation between the observations (so, if an observation is more correlated with the observation to be predicted, it gets a larger weight). Generally, but not always, locations that are closer in space are more correlated, and so observations that are closest to the location to be predicted get higher weight. However, outside of mining applications, covariation may depend on other non-Euclidean distance metrics (e.g., stream distance, city-block distance, etc.) or may even depend on observations relatively far away.

Although the initial implementations of Kriging did not include measurement (observation) error, nor underlying trend surfaces, this so-called "Universal Kriging" is now the standard approach and is similar to linear mixed model spatial regression with dependent errors (see Cressie 1993 for a comprehensive overview). More generally, optimal spatial prediction is now considered in terms of Gaussian processes, which gives it much the same predictive power as many modern machine learning methods. In addition, spatial prediction is often considered for non-Gaussian data as well, analogously to generalized linear mixed models (i.e., so-called "model-based geostatistics"; see Diggle et al. 1998, and Cressie and Wikle, 2011, for an overview). A major challenge in modern applications is dealing with high-dimensional datasets, multivariate processes, spatio-temporal processes, and the realistic specification of covariance structure. This remains a very active area of research.

- Cressie, N. (1990) The origins of Kriging. Mathematical Geology, 22:239-252.
- Cressie, N.A.C. (1993). Statistics for Spatial Data, revised edition. John Wiley & Sons, Hoboken.
- Cressie, N. and Wikle, C.K. (2011). Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken.
- Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model-Based Geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(3), 299-350.
- Matheron, G. (1962). Traité de géostatistique appliquée. Editions Technip.

Abbreviated Background Notes on Small Area Estimation

Small Area Estimation (SAE) seeks to produce reliable estimates of characteristics for units (small geographic areas and/or subpopulations) when sample sizes are very small or even zero. A very extensive literature chronicles decades of work on these problems, including assessment of precision and uncertainty intervals.

Rao and Molina (2015), Small Area Estimation, 2nd ed. published by Wiley is a standard reference. Pfefferman's review article in Statistical Science (2013) summarizes both the problem and advances up to that time.

The "small" in the term SAE applies to the size of the sample in the area, not to the size of the area itself. The "area" in the term SAE may be defined geographically, but equally may be defined by domains or by other classifications such as socio-demographics. What is at the core of SAE problems is that point estimates and error measures are required for every area separately, not just as an average over some or all of the areas under consideration. In consequence, SAE must provide predictions for areas (or units) with no sample observations as well as estimates for areas with few observations. Hence SAE depends on relevant variables and requires some way of aggregating (pooling) like units or areas.

For small domain inference, there are basically two types of model: unit and area level models. The distinction is important. Area level models are based on aggregated direct survey estimates and area level covariates. By contrast, unit level models are based on observations from individual units and covariates observed at both the unit and area levels.

The basic area level model is the Fay-Herriot model (see Rao and Molina 2015). Since area level models are specified only at the area level, inferences are essentially about the model parameters, not about the set of individual non-sampled units. However, this model is used to improve precision when direct estimates are not precise enough for inference for small domain parameters.

By contrast, unit level models have more complex structure that is needed to produce predictions for non-sampled units.

Issues of selection bias (for example, when subpopulations are oversampled) are magnified when data sources with different selection probabilities are integrated - and this is a concern for all methods. Also, the choice of covariates is crucial because all methods depend on a set of covariates with good predictive power for the small area quantities of interest.

There are a variety of specific methodologies both Bayesian and frequentist that have been developed for specific applications with different data requirements and availability at unit and area levels and different quantities (means, quantiles, ordered means, etc.) to be estimated. Also, mapping (e.g., poverty mapping, disease mapping, etc.) has been done at both area and unit levels, depending on the kind of inference to be drawn.

Appendix C: List of Participants

Technical Experts

Bettina Aten, Bureau of Economic Analysis

Thesia Garner, Bureau of Labor Statistics

Dan Goldhaber, University of Washington-Seattle

Robert Hauser, American Philosophical Society

Helen Ladd, Duke University

Marguerita Roza, Georgetown University

Chris Wikle, University of Missouri

NISS

Nell Sedransk

Alexandra Brown

NCES

Peggy Carr

Chris Chapman

James Deaton

Douglas Gerverdt

Sarah Grady

Tracy Hunt-White

Joel McFarland

Jennifer Nielsen

Laura Nixon

Ross Santy

Marilyn Seastrom

Tom Snyder

William Ward

Andrew White

James “Lynn” Woodworth

Appendix D: Biosketches of Technical Experts

***Bettina Aten**, Ph.D.*

Title: Economist, Bureau of Economic Analysis, USDOL

Bettina Aten is Chief of Regional Prices at the Bureau of Economic Analysis (BEA), United States Commerce Department. Her recent work focuses on estimating geographic differences in price levels for states and metropolitan areas in the US. Geographic price levels are similar to purchasing power parities, a concept that is used in international comparisons of national accounts such as those published in the Penn World Table, which Ms. Aten co-produced for many years. Prior to joining BEA in 2003, Ms. Aten taught at the University of Illinois Champaign-Urbana and at Bridgewater State College in Massachusetts. She has a Ph.D. in Regional Science from the University of Pennsylvania, MBA from the Wharton School and a Computer Science degree from the University of São Paulo, Brazil.

***Thesia I. Garner**, Ph.D.*

Title: Supervisory Research Economist, Bureau of Labor Statistics, USDOL

Thesia I. Garner is a supervisory research economist in the Division of Price and Index Number Research, Bureau of Labor Statistics, U.S. Department of Labor. She is currently serving on the Interagency Working Group (TWG) on Evaluating Alternative Poverty Measures, and in 2010, she also served on the ITWG on Developing a Supplemental Poverty Measure for the U.S. Throughout Dr. Garner's professional career, her primary focus has been on the measurement of economic well-being, conducting research with colleagues from the Census Bureau and other federal agencies, along with colleagues internationally. She has focused her work on economic inequality using expenditure and income data, valuing owner-occupied housing for macro and micro economic statistics, subjective measures of economic well-being, supplemental poverty measurement thresholds, and joint distributions of income, consumption and wealth, and has published her work in numerous professional journals. Dr. Garner was a Fulbright Scholar to the Czech and Slovak Federal Republics from 1992-94, has served on the governing Council of the International Association for Research on Income and Wealth (IARIW), and is currently the Book Editor for the Review of Income and Wealth (RIW). Dr. Garner has also served on Expert Panels for the International Labour Organization (ILO) and OECD. In 2016, she was awarded the Roger Herriot Award along with Kathleen Short for their work on the poverty measure. Thesia holds a Ph.D. from the University of Maryland, a MS from Purdue University, and a BA from Meredith College.

***Dan Goldhaber**, Ph.D.*

Title: Director, CALDER, American Institute for Research & Director, CEDR, University of Washington

Dr. Dan Goldhaber is the Director of the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research and the Director of the Center for Education Data & Research (CEDR) at the University of Washington. His work focuses on issues of educational productivity and reform at the K-12 level, the broad array of human capital policies that influence the composition, distribution, and quality of teachers in the workforce, and connections between students' K-12 experiences and postsecondary outcomes. Topics of published work in this area include studies of the stability of value-added measures of teachers, the effects of teacher qualifications and quality on student achievement, and the impact of teacher pay structure and licensure on the teacher labor market. His research has been regularly published in leading peer-reviewed economic and education journals such as: American Economic Review, Journal of Human Resources, Journal of Policy and Management, Economics of Education Review, Education Finance and Policy, and Educational Evaluation and Policy Analysis. The findings from these articles have been covered in more widely accessible media outlets such as National Public Radio, the New York Times, the Washington Post, USA Today, and Education Week.

***Robert Hauser**, Ph.D.*

Title: Executive Officer, American Philosophical Society

Robert M. Hauser is the Executive Officer of the American Philosophical Society. From 2010 to 2017, he served as Executive Director of the Division of Behavioral and Social Sciences and Education at the National Academies of Sciences, Engineering, and Medicine. He is the Vilas Research Professor and Samuel Stouffer Professor of Sociology, Emeritus, at the University of Wisconsin-Madison. He has been an investigator on the Wisconsin Longitudinal Study (WLS) since 1969 and led the study from 1980 to 2010. The WLS, which began as a study of post-secondary education, has followed the lives of more than 10,000 Wisconsin High School graduates of 1957 for 60 years and has become a national resource for bio-social research on health and retirement. While at the UW-Madison, he directed the Center for Demography of Health and Aging, the Institute for Research on Poverty, and the Center for Demography and Ecology. Hauser's research interests include statistical methodology, trends in social mobility and in educational progression and achievement, the uses of educational assessment as a policy tool, and changes in socioeconomic standing, cognition, health, and well-being across the life course.

***Helen F. Ladd**, Ph.D.*

Title: Professor Emerita, Duke University

Helen F. Ladd is the Susan B. King Professor Emerita of Public Policy and Economics at Duke University's Sanford School of Public Policy. Her education research focuses on school finance and accountability, teacher labor markets, school choice, and early childhood programs. With colleagues at Duke University and UNC, she has studied school segregation, teacher labor markets, teacher quality, charter schools, and early childhood programs. With her husband, Edward Fiske, she has written books and articles on education reform efforts in New Zealand, South Africa, the Netherlands, and England. She is the co-author or co-editor of 12 books. These include Holding Schools Accountable: Performance-Based Reform in Education (Brookings Institution, 1996); The Handbook of Research in Education Finance and Policy (2008 and second edition 2015), books on school reform in New Zealand and South Africa, and Educational Goods: Values, Evidence and Decision Making (University of Chicago Press, 2018). Her prior experience includes co-chair of a National Academy of Sciences Committee on Education Finance, membership in the National Academy of Education, president of the Association for Public Policy and Management and co-chair of the national campaign for a Broader, Bolder Approach to Education. She taught at Dartmouth College, Wellesley College, and at Harvard University, first in the City and Regional Planning Program and then in the Kennedy School of Government.

***Marguerite Roza**, Ph.D.*

Title: Director, Edunomics Lab, Georgetown University

Dr. Roza directs a research center focused on education finance called the Edunomics Lab (Edunomicslab.org). Recent research traces the effects of fiscal policies at the federal, state, and district levels for implications on resources at school and classroom levels. Her calculations of dollar implications and cost equivalent tradeoffs have prompted changes in education finance policy at all levels in the education system. She's led projects on state and school district finance policy, financial equity, pensions, compensation, higher education finance, and other related topics. She served as Senior Economic Advisor to the Bill and Melinda Gates Foundation. Her work has been published by the Brookings Institution, Public Budgeting and Finance, Education Next, the Journal of Public Finance and Management, and the Peabody Journal of Education. Dr. Roza is author of the highly regarded education finance book, Educational Economics: Where Do School Funds Go? Dr. Roza teaches as part of the McCourt School's Certificate in Education Finance, and in programs elsewhere including University of Washington's MEP, Rice University's REEP program, and the Broad Center's Leadership Program.

***Christopher K. Wikle**, Ph.D.*

Title: Professor, University of Missouri

Christopher K. Wikle is Curators' Distinguished Professor and Chair of Statistics at the University of Missouri (MU), with additional appointments in Soil, Environmental and Atmospheric Sciences and the Truman School of Public Affairs. He received a PhD co-major in Statistics and Atmospheric Science in 1996 from Iowa State University. He was research fellow at the National Center for Atmospheric Research from 1996-1998, after which he joined the MU Department of Statistics. His research interests are in spatio-temporal statistics applied to environmental, ecological, geophysical, agricultural and federal survey applications, with particular interest in dynamics. His work has been concerned with formulating computationally efficient deep hierarchical Bayesian models motivated by scientific principles, with more recent work at the interface of deep neural models in machine learning. Awards include elected Fellow of the American Statistical Association (ASA), elected Fellow of the International Statistical Institute (ISI), Distinguished Alumni Award from the College of Liberal Arts and Sciences at Iowa State University, ASA Environmental (ENVR) Section Distinguished Achievement Award, co-awardee 2017 ASA Statistical Partnership Among Academe, Industry, and Government (SPAIG) Award, the MU Chancellor's Award for Outstanding Research and Creative Activity in the Physical and Mathematical Sciences, the Outstanding Graduate Faculty Award, and Outstanding Undergraduate Research Mentor Award. His book Statistics for Spatio-Temporal Data (co-authored with Noel Cressie) was the 2011 PROSE Award winner for excellence in the Mathematics Category by the Association of American Publishers and the 2013 DeGroot Prize winner from the International Society for Bayesian Analysis. His latest book, Spatio-Temporal Statistics with R, with Andrew Zammit-Mangion and Noel Cressie, was published in 2019 and is free to download at spacetimewithR.org. He is Associate Editor for several journals and is one of six inaugural members of the Statistics Board of Reviewing Editors for Science.

***Linda J. Young**, Ph.D.*

Title: Chief Mathematical Statistician and Director of Research and Development, USDA's National Agricultural Statistics Service

Linda J. Young is Chief Mathematical Statistician and Director of Research and Development of USDA's National Agricultural Statistics Service. She oversees efforts to continually improve the methodology underpinning the Agency's collection and dissemination of data on every facet of U.S. agriculture. Prior to joining NASS, Dr. Young served on the faculties of three land grant universities: Oklahoma State University, University of Nebraska, and the University of Florida. She has three books and more than 100 publications in over 50 different journals, constituting a mixture of statistics and subject-matter journals. A major component of her work has been collaborative with researchers in the agricultural, ecological, and environmental sciences. She has been the editor of the Journal of Agricultural, Biological and Environmental Statistics. Dr. Young has served in a broad range of offices within the professional statistical societies, including President of the Eastern North American Region of the International Biometric Society, Vice-President of the American Statistical Association, Chair of the Committee of Presidents of Statistical Societies, and member of the National Institute of Statistical Science's Board of Directors. Dr. Young is a fellow of the American Statistical Association (ASA), a fellow of the American Association for the Advancement of Science (AAAS), and an elected member of the International Statistical Institute (ISI).

Working Session convened by National Institute of Statistical Sciences

***Nell Sedransk**, Ph.D.*

Title: Director, National Institute of Statistical Sciences-DC

Dr. Nell Sedransk is the Director of the National Institute of Statistical Sciences. She is an Elected Member of the International Statistical Institute, also Elected Fellow of the American Statistical Association. She is coauthor of three technical books; and her research in both statistical theory and application appears in more than 60 scientific papers in refereed journals. The areas of her technical expertise include: design of complex experiments, Bayesian inference, spatial statistics and topological foundations for statistical theory. She has applied her expertise in statistical design and analysis of complex experiments and observational studies to a wide range of applications from physiology and medicine to engineering and sensors to social science applications in multi-observer scoring to ethical designs for clinical trials.

***Alexandra Brown**, M.S.*

Title: Research Assistant, National Institute of Statistical Sciences-DC

Alexandra Brown is a Research Assistant at the National Institute of Statistical Sciences working under the direction of Dr. Nell Sedransk on projects in education research. She holds a MS degree in Economics and is currently a PhD candidate in Survey Methodology at the University of Maryland.