DATA NEEDS FOR AGRI-ENVIRONMENTAL POLICY MODELING AND ANALYSIS

October 15, 2007

Workshop Summary
By Dennis Aigner

This workshop was organized by ERS in order to bring together users of agricultural data and data providers, with the goal of assessing data resources in support of so-called "agri-environmental policy modeling and analysis".

Since the inception in 1985 of the Conservation Compliance provision of the Farm Security Act, researchers have been engaged in work aimed at understanding who participates in voluntary conservation programs and under what circumstances, and assessing program effectiveness. Isolating the unique effects of such programs involves analysis that accounts for the influence of other factors that impact farm-level decision-making and thus requires data on all such factors.

Future agricultural policies may focus more attention on specific "ecosystem services" provided by the agricultural sector and either redirect agricultural subsidies and payments for working farms in this way or develop environmental regulations aimed at, e.g., reducing soil erosion or minimizing pesticide and fertilizer runoff. In either case, there will be specific data required to support policy design, to evaluate program effectiveness, and to improve program performance. While it may not be possible to foresee all future policy options, it is clear that the bulk of data requirements will be farm-specific (e.g., location, physical characteristics) and will involve demographic and other information on the farmer and the farm household.

In the opening presentation, Marca Weinberg of ERS lays out a schematic for analyzing farm conservation programs and their impact on the environment that identifies data requirements in a generic way, one that is by-and-large applicable to the task of program design and analysis of program effectiveness for future policy options involving ecosystem services or environmental outcomes.

In the next presentation, Vince Breneman of ERS summarizes the various datasets used by ERS researchers, including the Agricultural Resources Management Survey (ARMS), FSA's Common Land Unit (CLU) data, the National Resources Inventory (NRI), the Ssurgo and Statsgo soils databases, NASS's cropland satellite data, and other datasets involving climate and precipitation, elevation, hydrology, and transportation. A major issue for purposes of policy design and program evaluation is that some data are collected at the field or farm level while other relevant data are collected at the county, state or other aggregated levels, which leads to scaling/weighting problems. Most of the time, datasets are not very useful unless they are integrated with other sources, which makes solving these problems of great importance. Another major challenge concerns the fact that some datasets refer to complete coverage of the underlying observational units while

others reflect sampling of one form or another, posing yet another challenge for data "integration". Finally, these datasets are often controlled by different agencies, leading to another challenge: coordination.

Along these lines, Glenn Schaible of ERS discusses a recent effort at data integration: CEAP-ARMS. The Conservation Effects Assessment Program (CEAP) is a field-level survey of conservation practices and program participation corresponding to a physical location identified in the NRI. ARMS is a crop-specific, field/farm level survey of production practices, costs of production, farm finances, and farm resource and operator characteristics. These have been blended in CEAP-ARMS to allow for onsite environmental data to be taken into consideration when evaluating producer conservation behavior. This 2-year pilot project forcefully demonstrates the value of integrated data and successful cooperation between data agencies (ERS and NRCS).

The next two presentations, by Roger Claasen and Michael Hand of ERS, summarize efforts to analyze conservation programs by investigating aspects of producer participation, the producer decision to cultivate acreage in specific crops, and the resulting environmental consequences of land use decisions. In Claasen's work, the program/policy context is the existing one, consisting of crop insurance, the Conservation Reserve Program (CRP), and farm commodity payments. For Hand, the goal is to analyze a proposed policy aimed at reserving a certain amount of conservation assistance for beginning farmers and farmers who qualify as "socially disadvantaged", in this case so-called limited-resource producers. In both instances, the presenters review data limitations and aspects of data integration/aggregation, and provide their assessments of data "gaps".

The next three presentations assess data availability, data adequacy, and data "gaps" in the context of specific applications.

Rich Iovanna of FSA discusses the state of data available to support the analysis of the CRP, in particular, which fields ought to be in the CRP based on their potential to mitigate soil erosion, degraded water quality, and loss of wildlife habitat, compared to the fields farmers actually decide to put into CRP. In this case, FSA has lots of data but it's poorly organized. There are some inherent "structural" problems but mainly it's a matter of allocating sufficient resources to the tasks of data collection/construction (in particular, for water quality and wildlife habitat) and integration.

Scott Malcolm of ERS discusses data needs in the context of the Regional Environmental and Agricultural Policy (REAP) model, a major ERS effort to model crop rotation and tillage choices on U.S. farms that contains some relevant environmental outcomes, such as fertilizer and pesticide runoff. His main data issues revolve around the need to aggregate different datasets on a comparable regional basis. REAP could be implemented at a county or on a watershed basis, which would provide a better foundation for evaluating future agricultural policies focused at the farm level.

Bruce McCarl goes through a detailed analysis of data needs to support models that address three policy issues of current and future relevance: greenhouse gas emissions, the effects of climate change on cropping patterns, and animal biosecurity. McCarl's lists of data needed for these purposes are not superimposed on lists of agricultural data from its main purveyors, NRCS and NASS, and so the data "gaps" are not identified.

Though contained in the final section of the workshop, Jay Atwood's presentation of the EPIC and APEX simulation models really belongs among the data "users". The EPIC and APEX models simulate soil-type specific water movement and water-borne effluents (including nutrients, sediments, carbon, and pesticides) for homogeneous NRI areas (EPIC) or heterogeneous fields, farms, or watersheds (APEX). The models rely on farm survey data (NASS), NRI sample points, USDA field office records, and historical weather information. Atwood discusses three examples of application: CEAP current conditions, the National Nutrient Loss and Soil Carbon Database, and a scenario that asks whether CRP lands can be safely converted to corn production in light of governmental policies that favor domestic ethanol production.

The final section of the workshop was devoted to presentations by data providers.

The first of these, presented by Bob Kellogg of NRCS, concerns a survey designed around NRI sample points linking data on farm activities and conservation practices with NRI physical information in order to estimate the effects of conservation practices on the landscape. The current effort is aimed at national and regional estimation, but the data are being re-organized at the watershed level. The existing survey covers 2003-06. An update is planned to begin in 2011.

Stacy Wills and Doug Farmer of NASS provide an overview of NASS surveys and datasets that pertain to conservation programs and practices. These include ARMS, the Agricultural Labor Survey, various commodity surveys, the Census of Agriculture, and several Agricultural Chemical Use Surveys. Of specific relevance for agri-environmental policy and modeling are ARMS, the Agricultural Labor and Chemical Use Surveys, and the Census of Agriculture.

Lastly, Jeff Goebel of NRCS gives a short history of the NRI, noting in particular that its long history and longitudinal character makes it a unique resource for answering questions about land use changes and trends in soils, water, and other resources. Begun in 1982, NRI provides information on approximately 800,000 sample sites every five years through 1997, and annually since 2000 on a rotating subsample basis.

The wrap-up session of the workshop involved discussion by Otto Doering (Purdue University), John Stierna (American Farmland Trust), Jerry Fletcher (West Virginia University), and Dennis Aigner (University of California, Irvine). The main points raised revolved around the need to be forward-thinking in designing databases that can be helpful in addressing future policy scenarios, especially now when both agriculture and the policy environment seem to be in great flux. Also, since data for policy analysis often requires aggregation ("layering") or disaggregation, more satisfactory approaches to these

challenges need to be developed. Putting bounds of uncertainty on estimated policy outcomes is imperative, not only to reflect statistical confidence properly, but as a basis for determining how margins of error can be reduced by investing more in data collection, data quality, etc. In the case of simulated outcomes, similarly there should be a requirement to accompany predictions with the results of sensitivity analysis. A final point concerns confidentiality requirements and the legal limitations involved in working with data at the individual farm level, an issue that impedes current research in projects like CEAP-ARMS and may pose an impossible hurdle to overcome in pushing for even more farm-specific data. Thus, some effort needs to be expended in order to determine the absolute maximum level of disaggregation possible given legal limitations, and then to evaluate to what extent that hampers or constrains forward-looking policy design and analysis.