Econometric-Process Models of Semi-Subsistence Agricultural Systems: An Application of the Nutrient Monitoring Data for Machakos, Kenya

John M. Antle,[a] Gerdien W. Meijerink,[b] Andre de Jager,[b]
Jetse J. Stoorvogel,[c] and Alejandra M. Vallejo[c]

[a] *Department of Agricultural Economics and Economics, Montana State University, USA*

[b] *Agricultural Economics Research Institute (LEI), Wageningen University, The Netherlands*

[c] *Soil Science Centre/Lab of Soil Science and Geology, Wageningen University, The Netherlands*

June 2005

In *Tradeoff Analysis: An Operational and Accepted Tool for Policy Analysis,* a report prepared for the Ecoregional Fund for Methodological Initiatives.

http://www.tradeoffs.montana.edu/pdf/EP-Models.pdf

Econometric-Process Models of Semi-Subsistence Agricultural Systems: An Application of the Nutrient Monitoring Data for Machakos, Kenya

John M. Antle
Gerdien W. Meijerink
Andre de Jager
Jetse J. Stoorvogel
Alejandra M. Vallejo

June 2005

Semi-subsistence agriculture remains the dominant type of agriculture in the poorest and most environmentally vulnerable regions of the world. . Therefore, for quantitative analysis of poverty and sustainability, it is important to be able to characterize these systems in both bio-physical and economic terms. Doing so presents a challenge in terms of both data and modeling methodology. Significant advances have been made in developing computer-based systems for data collection. The Nutrient Monitoring system provides a systematic and comprehensive approach to characterize and collect data for both inputs into and outputs from agricultural systems. Once detailed data are available to characterize agricultural systems quantitatively, research administrators and policy analysts need modeling tools capable of assessing the behavior of these systems over time. Semi-subsistence agricultural systems exhibit a number of characteristics that make modeling them more difficult than systems typical of more commercially-oriented agriculture. Among these features are:

- a low degree of specialization and a high degree of diversification, with mixed crop-livestock systems common and a large number of different types of annual and perennial crops;
- inter-cropping (planting two or more species within an individual parcel of land);
- high rates of crop failure;
- extremely small field size and seasonal reconfiguration of sub-parcels within fields;
- limited use of purchased inputs, with positive amounts applied to some crops by some farmers (typically, marketed crops) and zero amounts applied to many subsistence (non-marketed) crops;

- high transportation and other transaction costs for purchased inputs and marketed outputs, and a lack of formal markets for some inputs and outputs;
- production credit available only through informal sources or unavailable.

The early applications of the econometric-process simulation modeling approach were to relatively simple systems, such as the potato-pasture system in Ecuador (Crissman, Antle and Capalbo, 1998) and the dryland grain production system in the US Great Plains (Antle and Capalbo, 2001). These models were formulated at the field scale to investigate agriculture-environment interactions such as pesticide leaching, soil erosion, and carbon sequestration. However, analysis of sustainability of semi-subsistence systems calls for a whole-farm approach, both to capture essential interactions between crops and livestock, and to assess measures of human welfare such as poverty and food security. Accordingly, the goal of this paper is to present an econometric-process simulation model for the semi-subsistence agricultural system found in the Machakos district of Kenya, using the Nutrient Monitoring data available for that system. The agricultural system in Machakos exhibits all of the characteristics listed above that are typical of semi-subsistence systems. We use the Machakos case to illustrate methods that can be used to construct a model that incorporates these characteristics, and also discuss limitations of these methods.

This paper begins with a brief review of the econometric-process simulation modeling approach (Antle and Capalbo, 2001), and its linkage to spatially-referenced data and biophysical process models using the Tradeoff Analysis software (Stoorvogel et al., 2004). The third section of the paper describes the Machakos production system using the Nutrient Monitoring data. The fourth section describes the development of the econometric-process model for Machakos. The fifth section uses this model to assess the potential for technology and policy interventions to impact poverty and sustainability of the system. The final section summarizes the paper, and discusses the strengths and limitations of the proposed modeling methods.

Econometric-Process Simulation Models and Tradeoff Analysis

As described in Stoorvogel et al. (2004), the basic concept underlying the econometric-process model approach is to estimate behavioral equations from conventional econometric production models for each activity in the system, and to then incorporate these behavioral equations into a simulation model that represents the structure of the farmer's decision making process on the extensive and intensive margins. The behavioral equations in the economic models are functions of conventional economic variables (output and input prices) and also estimates of inherent productivity derived from biophysical crop and livestock models. In this way, the economic models are linked to underlying biophysical conditions (soils, climate) that affect farmers' productivity and behavior. The extensive margin decisions of the farmer are represented as discrete land use decisions for each land unit. Once the farmer makes the decision about what to do on a land unit for a production cycle, the corresponding intensive margin (input-use) decisions are simulated. Given a sequence of land use and management decisions at a site (i.e., a farmer's field), any relevant environmental processes associated with those land use and management decisions are simulated. By simulating the land use and management decisions for a statistically representative sample of land units and decision makers, the results of this simulation exercise can be used to characterize the joint spatial distribution of these decision in a geographic region. These decisions can then be linked to environmental process models, such as the Nutrient Monitoring model, to construct a spatial distribution of the environmental consequences of these decisions. The resulting spatial distributions of both economic and environmental outcomes can then be statistically represented in maps and aggregated to derive tradeoff curves that provide information to support informed policy decision making.

Figure 1 shows the structure of the econometric-process simulation model for the crop-livestock system of Machakos. In this model, the variables that define a farm are location, farm size, family size, and number of livestock units. Spatial distributions for these variables are estimated, and these distributions are sampled to define the farms in a simulation. The parameters of these distributions, along with the parameters of the crop simulation models and the economic models, are the parameters that are used to define policy and technology scenarios with the econometric-process simulation model.

Characterizing Heterogeneity

A key feature of econometric-process simulation models is the heterogeneity of the populations being represented. Several aspects of heterogeneity enter an econometric-process model: the bio-physical heterogeneity of the land; the characteristics of the decision maker; and the economic heterogeneity associated with the location of the site. This heterogeneity is illustrated in Table 1 for Machakos. These data show the high degree of heterogeneity in terms of input use and crop use at the parcel level, and in terms of farm characteristics.

Bio-Physical Heterogeneity

Agricultural economists have devised a number of methods to capture the effects of soils, climate and genetic characteristics of crops in empirical economic models.

Many econometric models have been estimated with farm-specific or region-specific spatial dummy variables to capture spatial differences in productivity. Also many economists have included measurements of soil quality and climate in econometric production models. While these techniques may well capture effects of bio-physical characteristics on behavior, they fail to incorporate the systematic knowledge of the agricultural sciences about the relationships between the physical environment, genetic properties of crops, and crop productivity. Much of this knowledge has been embedded in modern crop growth models such as the DSSAT models (International Consortium for Agricultural Systems Analysis, 2004).

We have developed a procedure to systematically link bio-physical crop and livestock simulation models to econometric-process simulation models. The basic idea is that farmers base management decisions, in part, on their site-specific knowledge of production potential. We interpret the production predictions of bio-physical crop and livestock simulation models, specified with site-specific soils and climate data and average or representative management inputs, as a proxy for the farmer's knowledge about the spatial variation in productivity across sites. Thus, we use the bio-physical models' yield predictions as an index of productivity potential, not as a prediction of actual yield. To differentiate this variable from yield, we refer to it as a measure of a site's *inherent productivity*. We use inherent productivity variables as exogenous predictors of behavior in the estimation of econometric production models. When these econometric production models are then incorporated into the econometric-process simulation model, the underlying data and parameters that drive the bio-physical production models are also embedded in the economic model. The bio-physical models can be used to simulate out-of-sample behavior (e.g., behavioral response to environmental change, changes in genetic properties of crops) in ways that are consistent with the processes embedded in those models. This is not possible with statistical models alone. To implement this procedure, the Tradeoff Analysis software allows the user to run the DSSAT crop models and link them to econometric model estimation and simulation (Stoorvogel et al., 2004).[1]

To formalize this idea, let a production function be written in the general form $q = f(\mathbf{x}, \mathbf{z}, \mathbf{e})$ where $\mathbf{x}$ is a vector of variable inputs, $\mathbf{z}$ is a vector of fixed inputs, and $\mathbf{e}$ is a vector of bio-physical factors. Theoretically, soil and climate conditions define the potential productivity of a location that, combined with a plant type, management practices, and weather conditions, leads to a realized output. Crop growth simulation models can be represented in stylized form as $q = g(\mathbf{x}, \mathbf{e}, \gamma)$, for management $\mathbf{x}$, environmental variables $\mathbf{e}$, and genetic coefficients γ. Defining average or expected input use in the population as $\mathbf{x}^*$, we can use the crop growth simulation to calculate the inherent productivity q^* for a specific location on the basis of soil and weather data as $q^* = g(\mathbf{x}^*, \mathbf{e}, \gamma)$. As an alternative to the general model $q = f(\mathbf{x}, \mathbf{z}, \mathbf{e})$, we can specify the production function $q = h(\mathbf{x}, \mathbf{z}, q^*)$. Substituting for q^* we obtain $q = h(\mathbf{x}, \mathbf{z}, g(\mathbf{x}^*, \mathbf{e}, \gamma))$, showing that this procedure yields a special case of the general production function in which the variables $\mathbf{e}$ and γ are weakly separable from the variable and fixed inputs $\mathbf{x}$ and $\mathbf{z}$. This separability assumption is a testable hypothesis, given suitable data.

[1] We do not advocate using crop growth simulation models to predict the yields that are used to calculate economic returns in an economic model. There are a number of reasons why crop models may not provide realistic predictions of yield. For example, most crop growth models do not incorporate effects of pests. We use the crop growth models to integrate soils and climate data into a yield variable that we interpret as an index of productivity potential at a site, not as a realistic prediction of yield. In the econometric-process modeling approach, these simulated crop yields (or inherent productivities) are used along with prices and management decisions to predict observed output.

Our experience with this procedure shows that inherent productivity variables do provide a statistically useful way to systematically incorporate soils, climate, and genetic information into economic production models. There are two key elements to the successful implementation of this approach. First, the research team must have site-specific soils and climate data and crop genetic coefficients. Most secondary soils and climate data are not collected with sufficient spatial resolution to capture field-specific variation, so if that level of spatial resolution is needed, the research team will have to collect that information along with behavioral (i.e., farm survey) data. Alternatively, inherent productivity variables can be estimated that approximate conditions in a locale (e.g., a village or an agro-ecozone). Obviously, this approach loses some of the spatial variation in field-level data. Our experience suggests that when the available data span a sufficiently heterogeneous region this approach can still provide valuable information to support estimation of econometric production models. However, in some cases this approach fails because there is simply not enough variation in the soils and climate data to produce variation in the inherent productivity variable that is correlated with observed behavior. In such cases, it may be necessary to specify *a priori* the parameter linking inherent productivity to observed yield and then subject this parameter to sensitivity analysis (e.g., Gray, 2005).

While we have found this procedure to be a useful way to link bio-physical simulation models to econometric-process models, a number of unanswered questions remain about this approach. First, observe that the logic of inherent productivity presented above is based on the assumption that a single crop is grown with well-defined genetic properties. This assumption is appropriate for mono-cropped systems typical of agriculture in industrialized countries. But in many cases, notably the case of semi-subsistence farming in developing countries, farmers grow multiple crop varieties adapted to site-specific soil and climate conditions. We encountered this problem in our studies of the potato-based production system in the Peruvian Andes (see publications and related information at www.tradeoffs.montana.edu) as well as in the Machakos case. In the Peruvian system, farmers grow a number of traditional potato varieties and related tubers, primarily for home consumption, as well as improved varieties that are sold primarily for consumption in urban markets. A major production constraint at high altitude in the Andes is frost. The traditional varieties are more frost resistant than the improved varieties, yet many farmers still grow improved varieties at high altitudes because their market price is higher. In Machakos, farmers grow maize as a cash and subsistence crop, and also grow a complex mix of other crops, often in a single land parcel. Below we discuss some of the procedures we have employed to deal with this problem in the Machkos case.

Second, the discussion of inherent productivity thus far has been in a simple one-period model. In many situations, present management decisions influence future production, notably through effects on soil productivity. This raises the issue of system dynamics discussed in Antle and Stoorvogel (2005).

Decision Maker Heterogeneity

There is a large empirical literature showing that socio-economic characteristics of farm decision makers and farm households influence production decision making. One strand of the literature uses various measures of human capital (experience, years of schooling) as explanatory variables, beginning with the early work of Griliches and others in the 1960s (Huffman, 2001). A second strand of literature introduced farmers' risk attitudes as a factor influencing decision making, and researchers have hypothesized that attitudes toward risk and other attitudes that affect decision making vary across farmers, perhaps systematically with wealth, education, experience and other personal characteristics (Sunding and Zilberman, 2001). Another strand of literature is based on the household production model, wherein production decisions are modeled as non-separable from other household decisions (Strauss and Thomas, 1995). According to this approach, in principle any feature of the farm household (e.g., family size and composition, demographic characteristics, financial characteristics, etc.) could impact farm production decisions.

In principle, all of these features could be incorporated into an econometric-process simulation model, but in practice the ability to do so is limited by data availability, time and other resource constraints of the research team. Moreover, the point made earlier about modeling risk attitudes applies to most other farmer-specific or farm-household-specific characteristics. In many cases we can embed the parameters of the distributions of these characteristics into the reduced form parameters of the production model, and the model can be analyzed conditional on the underlying distribution of those characteristics in the population. Thus, in the Machakos model presented in Figure 1, the model uses location, farm size, family size, and livestock as characteristics of the farm household. We simulate each farm household's production system by sampling from the observed distributions of these characteristics in the population. Additionally, any other observable feature of the farm household that is a statistically significant predictor of behavior (e.g., gender of the farm decision maker, family composition, etc.) could be used to model the underlying spatial distribution of behavior in the population. In simulating the system, the parameters of the distributions of these exogenous variables (e.g., the means or variances) could then be manipulated to assess their impacts on the behavior of the system.

Economic Heterogeneity

The conceptual production model presented above is referenced by location. Clearly, prices faced by farmers vary spatially, so to construct a spatially-explicit economic simulation model we need to characterize the spatial distributions of prices. To implement the simulation model, each decision period we sample from these price distributions to represent the idea that each period farmers face randomly varying prices with systematic spatial components.

Price distributions can be modeled a number of ways, and it is beyond the scope of this paper to go into a methodological discussion. In our modeling work to date, we have used relatively simply recursive regression models to construct spatial distributions that reflect spatial correlations among farm-level prices in the models. In the literature on modeling spatial price distributions, the main focus of work is on hypothesis testing rather than developing models for simulation. Additional research is needed to investigate how best to model spatial price distributions so that they can be incorporated into simulation models.

Dynamics

Agricultural production takes place over time, with different time steps governing different interacting components. Agricultural systems are complex in the sense that they involve a number of interacting sub-systems. We have found it useful to categorize dynamics in terms of the way farmers make decisions. Intra-seasonal dynamics involve the sequence of decisions within a growing season. Intra-seasonal dynamics are most important in intensively managed systems, as exemplified by pest management decisions in systems such as the potato-pasture system of Ecuador in which large numbers of sequential pesticide applications are made (Crissman, Antle and Capalbo, 1998). However, in less-intensively managed systems (for example, the dryland small grain systems in the Great Plains of the United States, or semi-subsistence systems such as the one in Machakos), the inter-seasonal dynamics of crop rotations, the use of fallow, and interactions between crop and livestock systems are most important.

The interactions of crop and livestock systems with the environment are also inherently dynamic. The inter-seasonal dynamics of crop rotations are caused by dynamic interactions between management decisions and soil productivity. Thus, over time, the inherent productivity of the soil may change. A good example of this type of process is where farmers use the process of erosion to create slow-formation terraces and thus increase productivity of the system, as in Antle et al. (2005). More generally, the production of environmental services will be affected in the present period, and these environmental services may in turn affect future crop productivity or future environmental services. However, in the simplest characterization of the system, which we call *loose coupling*, each disciplinary component of the system may be dynamic but those dynamic properties are not integrated dynamically. The logic of the model is linear in the sense that inherent productivity is determined by exogenous bio-physical conditions, and economic decisions (e.g., land use or management) do not feed back to affect inherent productivity. Economic decisions may in turn affect environmental outcomes, but changes in environmental conditions do not feed back to inherent productivity or to economic decisions. In *close coupling*, the bio-physical and economic components of the model interact dynamically. This is clearly the case when management decisions impact soil productivity, and soil productivity in turn affects management decisions. Clearly, a major challenge facing the modeling of agricultural systems is to better capture the productivity dynamics (Antle and Stoorvogel, 2005).

In the Machakos model, the crop and livestock components of the model interact dynamically, thus capturing the effects of nutrients being cycled through the system. Crop residues are harvested and used as livestock feed, and manure and other organic amendments are accumulated and used on crops in subsequent seasons. However, the model is loosely coupled to the bio-physical crop models, in the sense that there are no feedbacks from the economic models to the crop models. Thus, the crop models provide the basis for predicting the spatial patterns of productivity implied by the baseline soils and climate data, but the model is not able to track changes in soil fertility. Likewise, the analysis of nutrient balances provided by the NUTMON model is a static accounting of nutrient flows and is not able to predict the dynamic changes in soil nutrients in response to changes in land use and management. To achieve this more dynamic analysis, better models of nutrient dynamics are needed, and once available these models need to be dynamically linked to economic decision models.

Modeling Diversified Crop Systems

A high degree of specialization is typical of high-productivity agriculture in the industrialized countries. In contrast, most low-productivity, semi-subsistence agricultural systems are highly diversified because farmers are partly meeting their household consumption needs as well as producing some crops for the cash market. This situation presents a challenge for modeling in several dimensions. First, even with a large sample of farms, survey data are unbalanced in the sense that in each growing season, farmers grow different combinations of crops, and often there are too few degrees of freedom for statistical modeling of minor crops. Second, prices are often not reported for products produced for home consumption. Third, many farmers practice inter-cropping of two or more crops. In some cases, inter-cropping involves a fairly standard practice across farms (as is the case with the maize-bean intercrop in Machakos), but in other cases farmers may combine a wide array of different crops on one land parcel. For example, in Machakos, farmers often mix a variety of grain, legume, vegetable and tree crops together on a single parcel.

From a theoretical point of view, inter-cropping is an example of a joint production process (one input vector produces more than one output). In cases where adequate data are available, and where a sufficient large number of observations are available for the same crop combination, it may be possible to use conventional multiple-output models (e.g., a multi-output profit or cost function). However, in many cases there is likely to be inadequate degrees of freedom, and farmers may be using different combinations of outputs. In this situation, it will be necessary to aggregate outputs and inputs in some form. The approach utilized in the Machakos model is to estimate a supply function specified as $v = f(p, w, z)$, where v is the value of all of the outputs, p is a quantity-weighted index of output prices, w is a vector of input prices, and z is a vector of other exogenous variables. Note that in the single-product case, this model could be written as $pq = f(p, w, z)$, where $f(.)$ is the conventional, single-product supply function.

Therefore, in this type of model, the quantity $(\partial f/\partial p)(p/v)$ can be interpreted as one plus the elasticity of supply.

The Problem of Small Field Size

A problem related to the highly diversified pattern of production is the frequency of extremely small field sizes. For example, in the Machakos data, the data show that individual parcels of land range from 0.01 hectare to over 5 hectares, with over 20 percent of the parcels less than 0.05 hectare (Figure 2). These extremely small parcel sizes are likely to make it difficult for enumerators to obtain accurate measurements of inputs and outputs and to inflate the effects of measurement errors when variables are calculated on a per-hectare basis. For example, when yields are calculated (output per hectare), extremely large and often implausible values are obtained. Therefore, researchers must, as always, carefully scrutinize the data to avoid having biases introduced. Ideally, these measurement errors will be corrected through quality control in the data collection process. In cases where obvious errors cannot be corrected after the data are collected, observations with extreme outliers should be be deleted from the analysis, or truncated to plausible values.

Models with Zero Values for Inputs and Outputs

Conventional production models assume firms produce a single output, or produce positive quantities of a fixed set of outputs, and they assume positive quantities of inputs are used. Many agricultural production systems violate these assumptions. Land use decisions at the level of the individual decision unit (the farmer’s field) necessarily involves the choice among a set of discrete alternatives. Many inputs in agricultural systems are non-essential, i.e., zero quantities may be used. Most production systems involve multiple outputs, with different farms producing different combinations of outputs. Many of these features of production data have not been addressed in the economics literature on production modeling because aggregated data are used. Wen we do attempt to model production with data for individual firms or farms, these problems arise frequently, just as they do in the micro-econometrics literature on labor supply where many of the estimation techniques for discrete choice and truncated and censored data were developed. The lack of attention to these issues may be partly explained by agricultural economists' use of optimization models in which some of these issues do not pose significant methodological challenges. Nevertheless, it is surprising that there has not been more discussion of these issues in the agricultural economics literature where many micro-econometric models of production have been estimated statistically. It would appear that these problems have often been ignored or assumed away by researchers. We cannot ignore them when we are modeling agricultural environment interactions, because it is essential to model site-specific decisions, and because these decisions interact in important ways with environmental processes.

The econometrics literature contains a variety of methods for dealing with these problems, but typically from the perspective of efficient estimation and hypothesis testing, not simulation. The goal here, however, is to obtain practical methods that provide a reasonable way to represent farmer decision making so that it can be both estimated and simulated. Pursuit of this goal has led us to develop methods that often depart from conventional econometric practice, because conventional econometric methods are often impractical or inappropriate from a simulation perspective.

Discrete Land Use Decisions

The earlier discussion of the econometric-process modeling approach alluded to the idea that farmers make discrete choices among alternative land uses. Therefore, the relevant question is how to model discrete land use decisions so that they can then be simulated. The econometric-process modeling approach is based on estimation of each production system independently, typically using the number of observations available for each type of output. These models are then used to simulate the variables that enter the farmer's objective function (e.g., expected returns), and the discrete land use decision is simulated by choosing the outcome with the highest value of the objective.

From an estimation efficiency perspective, the procedure described above is clearly statistically inefficient. Why not jointly estimate all the production systems accounting for across-equation correlations? One answer is that typically, farms produce different combinations of outputs, therefore, estimation of a system of supply and input demand equations with data from a cross section (or panel) of field-level data would involve estimating an unbalanced model. Moreover, simulation of a model with a complex error structure can be very difficult. Alternatively, why not estimate a multinomial discrete choice model? Here we can say first that available multinomial discrete choice models impose restrictive distributional assumptions, but more importantly, simulating discrete choice models involves very difficult computational problems.

Zeros in Input Data

As noted above, many agricultural inputs are non-essential, meaning that output can be produced with zero values of those inputs. One can find various discussions of this problem in the literature discussing how to estimate production functions with zero inputs (e.g., Johnson and Rausser, 1971; Battese, 1997).

We see this problem typically in pest management, where farmers may choose to treat for a pest if it is observed to be severe enough to cause economic damage, and may not treat otherwise. Moreover, in the case of pesticide applications, farmers are typically making a series of applications over time (an example of what we called intra-seasonal dynamics above). It is surprising that in most econometric studies of pesticide productivity, pesticides are aggregated over the growing season into a single quantity, thus tending to eliminate the zero input problem but also ignoring the sequential aspect of the problem and the implied input endogeneity. There are also significant issues with pesticide measurement, due to differences in input quality, which are often ignored as well. One solution we have developed to this problem is to model both the quantity and timing of individual pesticide applications in a system of dynamics factor demand equations. This approach finesses the zero-input problem, and avoids aggregating across time, but requires highly detailed data that are often lacking (Antle, Capalbo and Crissman, 1994).

In semi-subsistence agriculture in developing countries, we often see farmers applying low and even zero rates of mineral fertilizers as well as pesticides. For example, in the Machakos data, almost 80 percent of the parcels of maize have zero mineral fertilizer applied (Figure 3). Other inputs such as hired labor and animal labor are also non-essential. In this case, input use may be constrained by input availability and the farmer's financial situation. For example, in studies we have conducted in Peru, Senegal and Kenya, we have found that semi-subsistence farmers only applied purchased inputs to cash crops, and even then often 50 percent or more of the farmers may not use mineral fertilizers. This situation suggests that a discrete-continuous model is appropriate for factor demand equations, wherein the decision to use is made first, then positive quantities are determined in those cases where the input is used. The Heckman and related discrete-continuous choice models may be appropriate ways to estimate input demand equations, and can be simulated using univariate normal probability distributions if the discrete-choice component is modeled with a probit model. However, observe that application of this procedure raises very complex estimation and simulation problems if a system of factor demand equations is jointly estimated. Therefore, in our applications we have utilized single-equation estimation.

The presence of zero input quantities also creates problems for estimation of production functions or supply functions, because most convenient functional forms do not allow non-essential inputs (note the quadratic model does allow non-essential inputs, but imposes other restrictions and is not parsimonious in parameters). One simple approach we have developed and used with some success is based on the observation that in many data sets where zero input use is prevalent, quantities of inputs are not measured accurately making production function estimation difficult, and input prices are also measured inaccurately and lack sufficient variation to estimate supply functions (e.g., Gray, 2005). However, these data sets typically do accurately indicate whether or not an input is used. Therefore, a production function or supply function can be estimated using dummy variables indicating input use and non-use (alternatively, the dummy variable may be defined as indicating input use above or below a positive threshold value). The dummy variable parameter (say, in a Cobb-Douglas model) can then be interpreted as indicating the average productivity of the input in the user group. When this model is simulated, we approximate the production response for that input by assuming a montonic-increasing output response from zero up to the mean of the positive input-using group, and a lower (or zero) output response beyond that point.[2]

Zeros in Output Data

In highly commercialized agriculture dominated by extreme specialization, many farmers produce the same output or the same mix of outputs (e.g., the corn-soybean farmer in the Midwestern U.S., or the small-grain farmer in the Great Plains of North America). Many of these farmers actually produce a mix of crops, including oilseeds, hay, and other minor feedgrains. In semi-subsistence agricultures, often a larger number of different crops and livestock will be grown, both for subsistence of the farm household, for cash sale, and for production of important by-products such as organic fertilizer (as in the Machkos system illustrated in Figure 1). In these various systems, a complete data set on farm outputs will often have zero values for two reasons. First, farmers produce different crop mixes, over both space and time, as part of crop rotations, to manage risk, and for other reasons. Second, crop failures occur, and are particularly important in some systems. For example, in the Machakos system, the principal food and cash crop, maize, has a crop failure rate that averges 26 percent for the sample and is over 50 percent in some villages (Figure 4).

Considering the prevalence of this phenomenon, the problem of modeling multiple output systems with unbalanced data has received surprisingly little attention in the literature. One exception is the paper by Huffman (1988) in which he proposed using

[2] This procedure may also help reduce input endogeneity problems, since the use decision is made *ex ante* whereas application rate decisions are often made during the production process.

a mixed discrete-continuous model for estimation. Weininger (2003) reviews the literature and proposes a way to adapt the translog cost function to this situation. However, the complexity of these models would make simulation problematic (note that translog models are not globally well-behaved, so they are not attractive for simulation outside the range of observed behavior). The solution we have developed, as we noted above, is to simply estimate each production system (i.e., each crop or livestock system) independently, and then combine them to simulate the farmer's discrete land use decisions as the maximization of a well-defined objective function. To account for crop failure in the Machkos system, we have utilized a probit model to predict crop success or failure, and then used the probability of crop success to compute the expected value of the returns in the simulated land use decision. For the case of maize, this model shows that crop failure is strongly (negatively) related to crop inherent productivity and purchased input use, as would be expected.

Modeling Policy and Technology Scenarios: Primal vs Dual and Model Coupling

A major goal of agricultural system modeling is to simulate the effects of changes in policies and changes in production technologies. We have encountered several issues in doing these simulations that have implications for the type of economic modeling approach that is used.

A first observation is that either primal or dual systems could be used. Pros and cons of primal versus dual have been discussed at various places in the literature. From the perspective of modeling policy scenarios, where the main focus is how changes in prices and other exogenous factors (e.g., land use restrictions) impact farmer decision making, a dual approach is useful because it represents the response to price changes.

However, the dual approach is problematic when the goal is to represent technology changes, because supply functions and input demand functions are reduced form equations that embed the technology parameters without identifying them. Thus, a case can be made for the primal approach for modeling technology changes.

A related problem arises in linking bio-physical simulation models with economic models for the analysis of technology changes. A change in crop variety can be simulated with a crop growth model through the genetic parameters of the model. However, it is less clear how that change in variety translates into changes in economic model parameters. One might argue that our approach discussed above of incorporating inherent productivities into the estimation of economic models would solve this problem, but that is only true under the separability implied by that model. If separability is not valid, then it may be necessary to more closely couple the bio-physical and economic models. Ultimately, researchers have to trade off model complexity and data availability versus model validity in deciding how best to proceed.

Minimum Data Modeling

I argued in the first section that researchers need to pay more attention to developing models that provide a good enough answer for policy analysis using the minimum data possible. We have pursued this objective in modeling the supply of environmental services by noting that, as shown in Figure 1, the key information needed to model the adoption of alternative practices is the spatial distribution of opportunity cost. As a first-order approximation, for given prices and technology, this spatial distribution can be approximated with means and variances of the returns to each activity and the covariance of returns. Existing secondary data can be used to estimate these statistical parameters, and then the supply curve for environmental services can be derived by simply sampling repeatedly from these empirical distributions and allocating land units to each use according to maximization of expected returns. In recent unpublished work, we tested this approach using data from a case study of carbon sequestration in agricultural soils. The analysis showed that the carbon supply curve derived from a more detailed econometric-process simulation model was bounded by supply curves from a minimum-data model based simply on means, variances, and covariances of returns. This finding encourages us to believe that the minimum data approach could be a useful tool for producing timely analysis to support policy decision making.

Again, there is a tradeoff between the amount of information that the model can produce and the amount of information needed to estimate it. This minimum-data strategy takes the spatial distributions of returns for the competing activities as givens, so it is not possible to simulate scenarios in which, say, the price of an input changes, thus shifting the distributions. To infer these shifts, a more detailed model with output supply and input demand functions would be needed. But in cases where the policy decision makers only need to know the supply of the environmental service, taking prices as given, then the minimum data approach could provide that answer with much lower data demands that would be needed to estimate a system of supply and demand equations for each activity.

References

Antle, J.M. and S.M. Capalbo (2001). "Econometric-Process Models for Integrated Assessment of Agricultural Production Systems." *American Journal of Agricultural Economics* 83(2):389-401.

Antle, J.M., S.M. Capalbo, and C.C. Crissman. (1994). "Econometric Production Models with Endogenous Input Timing: An Application to Ecuadorian Potato Production." *Journal of Agricultural and Resource Economics* 19(1):1-18.

Antle, J.M. and J.J. Stoorvogel. (2005) "Incorporating Systems Dynamics and Spatial Heterogeneity in Integrated Assessment of Agricultural Production Systems." *Environment and Development Economics,* in press.

Antle, J.M., R.O. Valdivia, C.C. Crissman, J.J. Stoorvogel, and D. Yanggen (2005) "Spatial Heterogeneity and Adoption of Soil Conservation Investments: Integrated Assessment of Slow Formation Terraces in the Andes." *Journal of International Agricultural Trade and Development.* 1: in press.

Battese, George E. (1997) "A Note on the Estimation of Cobb-Douglas Production Functions When Some Explanatory Variables Have Zero Values." *Journal of Agricultural Economics* 48, 250-252.

Crissman, C.C., J.M. Antle, and S.M. Capalbo, eds., *Economic, Environmental, and Health Tradeoffs in Agriculture: Pesticides and the Sustainability of Andean Potato Production*. Boston: Kluwer Academic Publishers, 1998.

Gray, K.M. (2005). Changing Soil Degradation Trends in Senegal with Carbon Sequestration Payments. M.S. thesis. Montana State University, Bozeman, Montana, USA. www.tradeoffs.montana.edu.

Huffman, W.E. (1988). “An Econometric Methodology for Multiple-Output Agricultural Technology: An Application of Endogenous Switching Models.” S.M. Capalbo and J.M. Antle, eds. Agricultural Productivity: Measurement and Explanation. Washington, D.C., Resources for the Future.

Huffman, W.E. (2001). “Human Capital: Education and Agriculture.” G. Rausser and B. Gardner, eds., Handbook of Agricultural Economics, Volume 1A. Amsterdam, Elsevier.

International Consortium for Agricultural Systems Analysis, 2004. www.icasa.net/dssat/.

Johnson, S. R., and G.C. Rausser. (1971) “Effects of Misspecifications of Log-Linear When Sample Values are Zero or Negative.” *American Journal of Agricultural Economics* 53: 120-24

Stoorvogel, J.J., J.M. Antle, and C.C. Crissman. (2004). “Trade-off Analysis in the Northern Andes to Study the Dynamics in Agricultural Land Use.” *Journal of Environmental Management* 72(1-2):23-33.

Stoorvogel, J.J., J.M. Antle, C.C. Crissman, and W. Bowen. (2004). “The Tradeoff Analysis Model: Integrated Bio-physical and Economic Modeling of Agricultural Production Systems.” *Agricultural Systems* 80(1):43-66.

Strauss, J. and D. Thomas. (1995). Empirical modeling of household and family decisions. J. Behrman and T.N. Srinivasan, eds. Handbook of Development Economics, Volume IIIA. Amsterdam, Elsevier.

Sunding, D. and D. Zilberman (2001). “The agricultural innovation process: research and technology adoption in a changing agricultural sector.” G. Rausser and B. Gardner, eds., Handbook of Agricultural Economics, Volume 1A. Amsterdam, Elsevier.Weninger, Q. (2003). “Estimating Multiproduct Costs When Some Outputs are Not Produced.” *Empirical Economics* 28: 753-65.

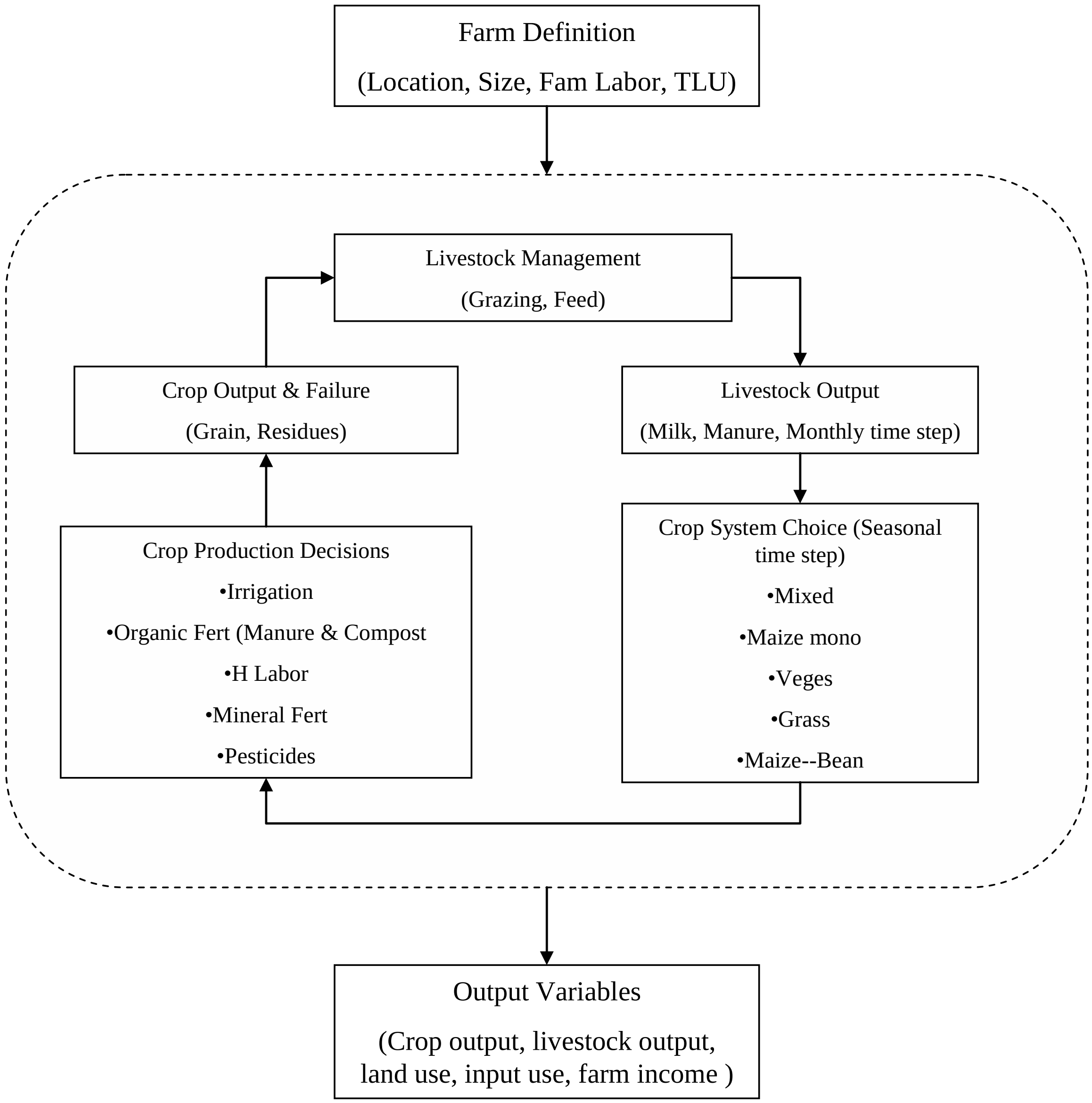

Figure 1. Structure of the Econometric-Process Simulation Model of the Crop-Livestock System in Machakos

Table 1. Parcel Means of Input Variables, System Frequencies and Farm Characteristicsin NUTMON Data for Machakos, by Village

	V1	V2	V3	V4	V5	V6
Inputs						
Parcel Size (ha)	0.35	0.23	0.56	0.98	0.10	0.55
Manure (kg/ha)	3996	231	582	347	828	185
Hired Labor (md/sea)	33	5	9	6	2	65
Mineral Fertilizer (kg/ha)	39	3	24	0	19	20
Seed (kg/ha)	57	54	54	21	57	15
Pesticides (kg/ha)	1.69	0.01	0.27	0.12	6.47	2.70
Systems						
Mixed Intercrop	0.48	0.18	0.23	0.16	0.38	0.53
Maize	0.36	0.09	0.16	0.20	0.22	0.08
Vegetables	0.28	0.00	0.05	0.00	0.16	0.26
Grass	0.43	0.17	0.16	0.10	0.03	0.09
Maize-Bean	0.38	0.55	0.41	0.54	0.21	0.04
Farm Characteristics						
Size (ha)	2.68	2.26	2.77	7.04	1.50	3.48
Livestock Units	1.30	1.94	1.65	2.22	2.15	1.08
Familly Labor (md/sea)	175	75	113	154	169	152
Literacy (%)	100	77	84	61	93	88
Famer Occupation (%)	46	50	52	64	82	76
Family Size	8.77	8.27	6.84	7.24	8.50	7.35
Off-Farm Income (ks/yr)	5255	2770	15078	7873	2405	8814

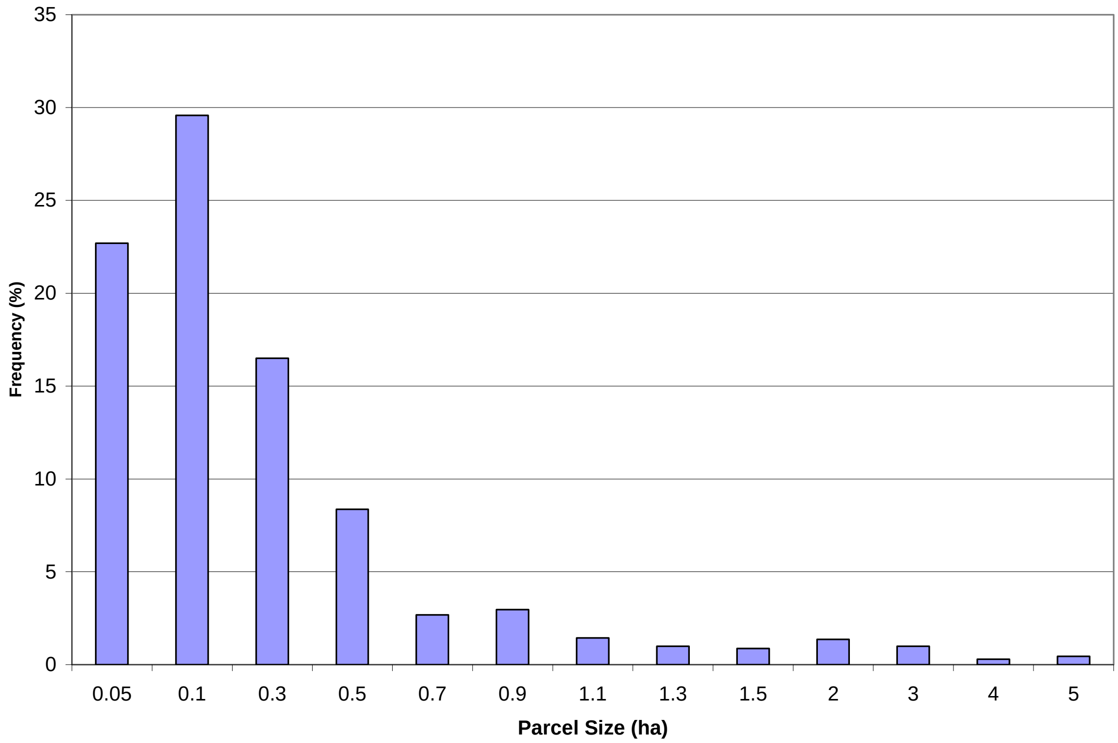

Figure 2. Parcel Size Distribution in Machakos

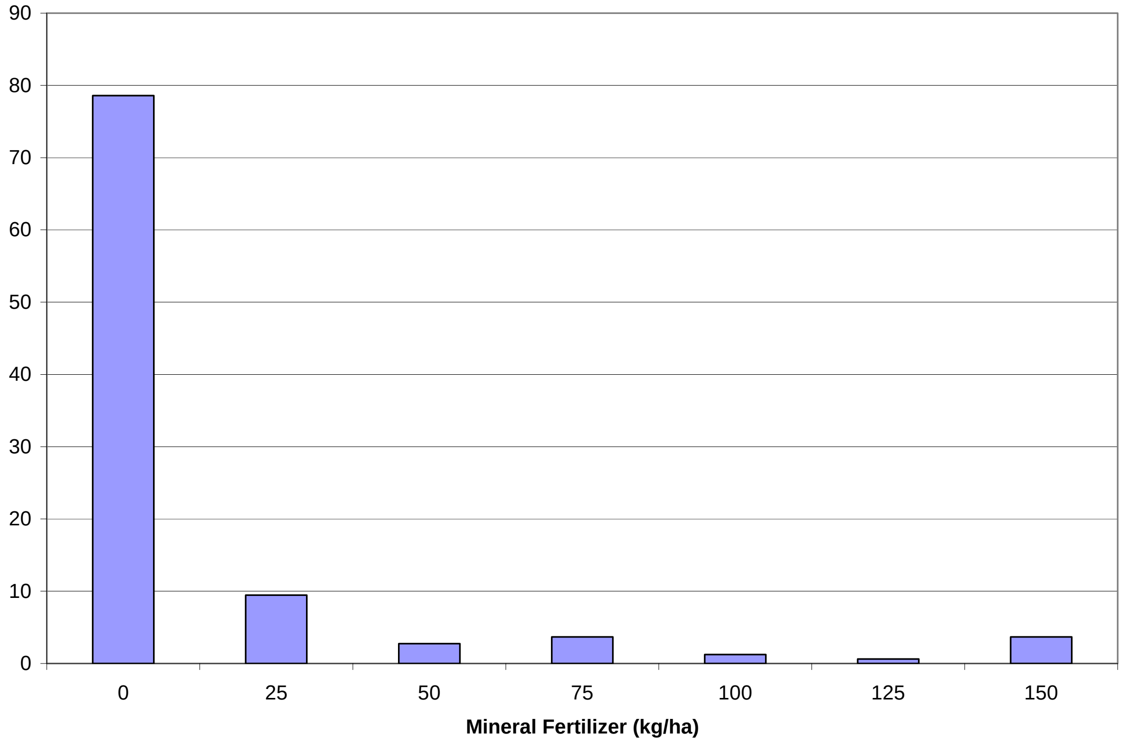

Figure 3. Frequency Distribution of Mineral Fertilizer Use in Maize, Machakos

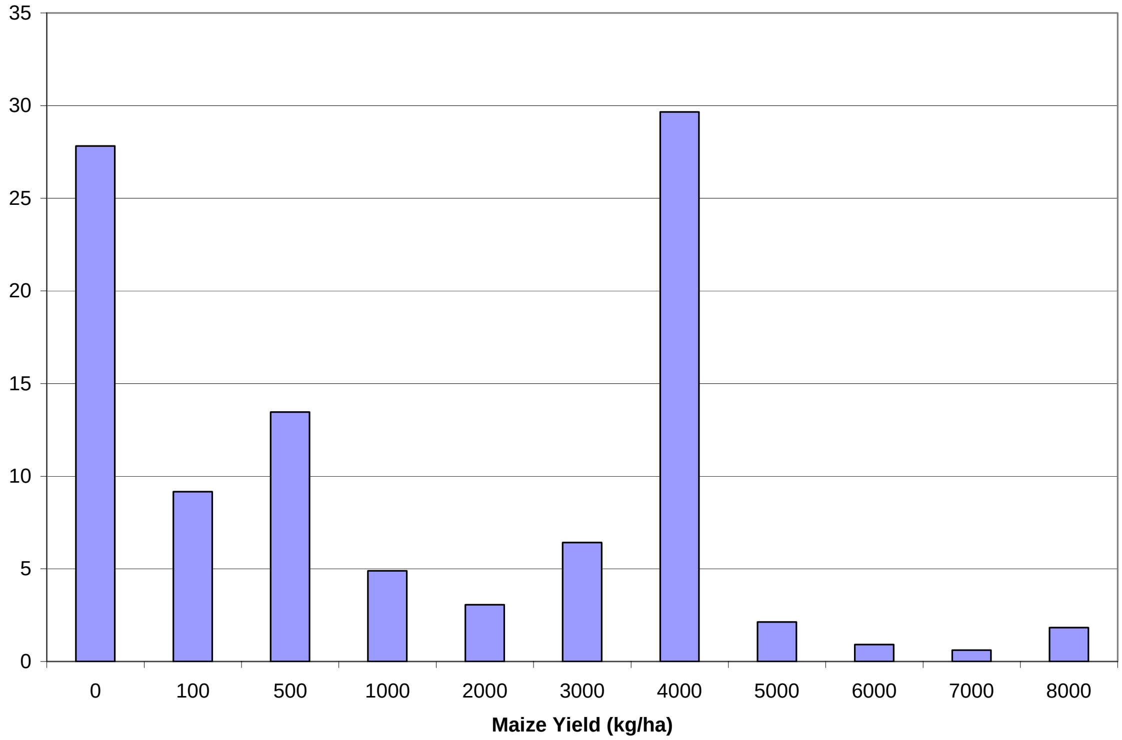

Figure 4. Maize Yield Distribution for Machakos