Reconstructing Disaggregate Production Functions

by
Richard E Howitt* and Siwa Msangi*

Selected paper to be presented at the 2002 AAEA-WAEA Annual Meeting
Long Beach, California, July 28-31, 2002

* Department of Agricultural and Resource Economics, University of California, Davis, California, 95616 email howitt@primal.ucdavis.edu

Introduction

This paper develops a method for reconstructing flexible form production functions using minimal disaggregated data sets. We use the term reconstruction rather than estimation because many disaggregate data sets have insufficient observations to claim statistical properties for the resulting parameter estimates. Since we are interested in models that can address policy questions, the emphasis in this paper is on the ability of the model to reproduce the existing production system and predict the disaggregated outcomes of policy changes.

In many developed and developing agricultural economies there is considerable emphasis on the effect of agricultural policies and production on the environment, and conversely, the effect of environmental policies on the agricultural sector. This emphasis is likely to lead to the rehabilitation of production function models for many policy problems, in the literature. There are several reasons why production functions are suited to the analysis of agricultural- environmental policy. First, environmental values are measured in terms of the physical outcomes of agricultural activity. Second, primal data on crop yields, areas and input use is usually more readily available and often more accurate than cost data. Third, constraints and subsidies on some inputs require that shadow prices be added to nominal prices. Moreover some environmental policies are formulated as constraints on input use. Fourth, economic models of agricultural and environmental policy impacts often have to formally interact with process models of the physical systems. Such models require the economic output in terms of primal values.

Several authors have emphasized the need to spatially disaggregate models for environmental policy analysis (Antle & Capalbo, 2001; Just & Antle, 1990). However, such disaggregation is often made difficult either by the limited availability of disaggregate data or, if such data is present, the lack of enough degrees of freedom to identify disaggregate parameters within a classical estimation framework. Generalized Maximum Entropy (GME) estimation techniques (Golan et al., 1996a) have come into increasing use by researchers who seek to achieve higher levels of disaggregation in the face of these data problems (Lence & Miller, 1998b); Lansink et al., 2001; Golan et al., 1994, 1996b). Given the inherent heterogeneity of soils and other agricultural resources, the researcher who wishes to disaggregate cross-sectional data must consider the trade-off between two possible sources of model error: that caused by aggregation bias versus that due to small sample bias. Aggregating across heterogeneous regions leads to aggregation bias, whereas ill-conditioned or ill-posed GME estimates may be biased due to the small sample on which they are based. An additional advantage that speaks in favor of maximum entropy-based alternatives is the ability to formally incorporate additional data or informative priors into the estimation process, in a Bayesian fashion.

Substitution activity at the intensive and extensive margins is a key focus of agricultural-environmental policy analysis. A basic policy tool is to provide incentives or penalties that lead to input substitution under the given agricultural technology. Such substitutions at the intensive margin can reduce the environmental cost of producing traditional agricultural products or that of jointly producing agricultural and environmental benefits. These policies cannot be evaluated without explicit representation of the agricultural production process. It follows, therefore, that the potential for substitution should be explicitly modeled within a multi-input multi-output production framework.

The reconstructed production function (RPF) modeling approach proposed in this paper can be implemented with smaller data sets than those required by conventional econometric approaches. However, RPF models still result in production functions that have all the properties enjoyed by flexible functional forms, such as multi-input, multi-output quadratic, square root, generalized Leontieff and trans-log specifications. This is achieved by using a combination of GME estimation and positive mathematical programming (PMP) calibration techniques (Howitt, 1995), so as to reconstruct a well-behaved production function that is entirely consistent with the data and which calibrates to the base conditions. This combination of methodologies distinguishes our approach with other GME production analyses in the literature (Zhang & Fan, 2001; Lence & Miller, 1998a). The GME estimates given in this paper do, however, converge to consistent estimates when the sample size is increased and have been shown to have the same asymptotic properties as conventional likelihood estimators (Mittlehammer et al., 2000). We have the assurance of consistency, in our approach, by the inclusion of moment constraints in the estimation procedure, which are absent from the formulations presented in other applications of GME in the literature – such as the aforementioned studies.

The ability to simulate policy alternatives reliably with constrained profit maximization requires calibration using the marginal and total product conditions and stability in the second-order profit-maximizing conditions. It is our belief that those who make use of policy models are more interested in both reproducing observed behavior and then simulating outcomes beyond the base scenario – than in testing for the curvature properties of the underlying production function. Therefore, the RPF models presented here are reconstructed subject to parameter restrictions that result in locally concave production functions. Within our programming-based reconstruction and simulation framework, we can also impose policy restrictions in the form of constraints on the reconstructed model.

Section II of the paper briefly reviews modeling methods used to estimate the effect on land use of agricultural and environmental policies. Section III develops the reconstruction procedure for the production model, within the GME framework. Section IV contains an empirical application that measures the production response of California's irrigated crop sector to environmentally-driven policy changes in irrigation water supplies. This is followed by Section V, in conclusion to the paper.

II. Methods for Modeling Disaggregated Agricultural Production and Land Use.

The approach that we use in this paper addresses the shortcomings of representative farmer models enumerated by Antle & Capalbo (2001), when they cite the limited range of response in the typical representative production model. The embedded PMP parameters capture the individual heterogeneity of the local production environment, be it in terms of land quality or other site-specific effects, and allow the estimated production function to replicate the input usage and outputs produced in the base year.

Love (1999) made the point that the level of disaggregation matters, in terms of the degree of firm-level heterogeneity and other localized idiosyncrasies that get averaged out of the sample. This affects the likelihood of observing positive results for tests of neo-classical behavior, such as cost minimization or profit maximization. In our approach, we impose curvature conditions on the reconstructed production function, since we are aiming for models that reproduce behavior rather than test for its theoretical properties. The relative stability we observe within cropping systems, despite the presence of substantial yield and price fluctuations is strong empirical evidence that farmers act as if their profit functions are convex in crop allocation. The gradual adjustment of agricultural systems to changes in relative crop profitability suggests that farmers adjust by progressive changes over time, along all the margins of substitution, rather than going from one extreme corner solution to the next.

Zhang & Fan (2001) conclude that the behavioral assumptions of profit maximization are too strong for the example to which they applied a GME production function estimation. While their level of aggregation was severe, they made the case for using GME on the basis of its ability to incorporate non-sample information and to deal with imperfectly observed activity-specific inputs. Within our framework, we are able to implement more flexible functional forms for production than that used by Zhang & Fan, as well as to avoid imposing constant returns to scale, as a result of our higher level of disaggregation.

Just et al (1983), stated in their classic production paper that their:

> "Methodology is based on the following assumptions that seem to characterize most agricultural production:
>
> *(a)Allocated inputs*. Most agricultural inputs are allocated by farmers to specific production activities..
>
> (b)*Physical constraints*. Physical constraints limit the total quantity of some inputs that a farmer can use in a given period of time …
>
> (c) *Output determination*. Output combinations are determined uniquely by the allocation of inputs to various production activities aside from random, uncontrollable forces."

Just et al's specification admits jointness in multioutput production only through the common restrictions on allocatable inputs. The specification in this paper has constraints on the available land, but also allows for jointness between crops in a given region, as reflected by the deviations of crop value marginal products from the opportunity cost of restricted land inputs.

The current range of approaches to agricultural production modeling and the associated analysis of environmental impacts, seems to fall into three groups; namely, disaggregated calibrated or constrained programming models (McCarl, 2000; Alig et al., 1998; CVPM[1], 1997; CAPRI[2], 2000) disaggregated logistic land use models (Wu & Babcock, 1999), and aggregate econometric land use models (Mendelsohn et al., 1994).

In this paper we hope to straddle the current divide between programming and econometric approaches to production analysis, not only by using constraints and output determination in our formulation, but also by using flexible functional forms and data in which the principle explanatory variables of yields, prices and crop land allocation are based on small stochastic samples.

[1] Central Valley Production Model , used in the 1997 Programmatic Environmental Impact Statement of the Central Valley Project Improvement Act (see references).

[2] Common Agricultural Policy Regional Impact (http://www.agp.uni-bonn.de/agpo/rsrch/capri/)

III Using Generalized Maximum Entropy to Reconstruct Production Functions

The term 'reconstruction' was developed in the field of image processing where the problem of reconstructing images from incomplete data is, by definition, ill-posed. Image reconstruction uses structural information about the image to generate a complete image from sparse or incomplete data observations. A simple example is to use curvature criteria to reconstruct an image of the rings around Saturn based on relatively few pixels sent from outer space. The analogy in production models is to reconstruct the production surface for a specific crop at a specific location by using a small set of observations of the average product and marginal input allocations.

Reconstruction methods are now used in several scientific fields, among which are tomography, astronomy, and the earth sciences. We view the problem facing agricultural policy modelers as one of trying to reconstruct the behavioral and technical relationships that drive agricultural production decisions, at a scale that has policy relevance for those environmental resources affected by agricultural activity. Often the scale of disaggregation that is required for meaningful environmental policy analysis, differs widely from that of the economic data set available. It follows, therefore, that a flexible form production function model on the same level of disaggregation as that desired for environmental policy will suffer from low degrees of freedom or be ill-posed. Hence the need for a reconstruction approach. Consequently, maximum entropy estimation is a commonly used basis for reconstruction algorithms in the physical sciences (Desmedt 1991), and is coming into increasing use in agricultural economics.

The nature of the data set defines the precise reconstruction method to be used. For disaggregated policy models, the available data usually takes the form of short time-series at the desired level of disaggregation, or a cross-sectional survey sample taken over each disaggregated region. A reassuring characteristic of generalized maximum entropy (GME) estimators is that, while they effectively estimate ill-conditioned or ill-posed problems, they are consistent in large sample, and enjoy the usual classical properties of estimators (Mittlehammer et al, 2000). The GME reconstruction approach advanced in this paper is completely in accord with classical econometric procedures for large sample problems. The novelty of the paper lies in the idea that the modeler does not have to accept the stricture of non-negative degrees of freedom, but may specify a complex model at the level of disaggregation that is thought to minimize the effect of estimation and aggregation bias on the model outcome. The modeler can specify flexible multi-input production functions for any number of observations and calibrate closely to the base conditions. Essentially we show that a minimal level of data, that would have restricted the modeler to a simple linear programming model, in the past, can now be used to calibrate and reconstruct a set of multi-input quadratic production functions. Other functional forms that have continuity and concavity can also used, namely the trans log and generalized Leontieff form. In the discussion below we use a quadratic production function, and model the multi-outputs from a disaggregated unit using the specification in Just et al. (1983).

The first-order conditions for optimal allocation have to incorporate the shadow value of any constraints on input usage. Since the allocatable inputs are restricted in quantity, and rotational interdependencies can exist between crops, we use a modified PMP model on each data sample to obtain a numerical value for a shadow price that may exist above or below the allocatable input cash price. Specifically, we impose upper- and lower-bound calibration constraints on the crop allocation in each sample to generate the additional economic values that may be observed due to rotational interdependencies or land heterogeneity[3]. For cases where the data actually contains reliable information on rental markets, the reconstruction proceeds directly, without the intervening PMP stage – whose only role is to generate shadow values for rotations and allocatable inputs that are consistent with the data.

Assume that we have "n" observations over time on a farm unit that produces "j" crops, each of which has "i" inputs. There is a subset of restricted, but allocatable inputs, such as land or irrigation water. The data set consists of n observations on crop price, input price, crop yield, and input use by crop. This data set and other agronomic data can be used to define the implicit Leontieff matrix A and specify the following calibrated linear programming problem.

$$(1) \qquad Max\, \Sigma_j \; p_j y_j x_j - \Sigma_{ij} \omega_j a_{ij} x_{ij}$$

$$\begin{aligned} s.t \qquad & Ax \leq \; b \\ & Ix \leq \tilde{x} + \varepsilon \\ & Ix \geq \tilde{x} - \varepsilon \\ & x_{ij} \geq 0 \end{aligned}$$

where b is the vector of available input quantities, p and ω are the output and input price vectors (respectively) and y is the yield per acre, and x is the decision variable for cultivated land area.

The first set of allocatable resource constraints generates the shadow values for those constraints that influence the observed crop and input allocations. The perturbed upper- and lower-bound calibration constraints ensure that the crop allocation is within ε of the observed data[4], and in addition, provide measures of the rotational cost interdependence between the crops based on the equi-marginal principle for land allocation via their shadow values.

A generic GME reconstruction problem can be written as

$$\max_{\bar{p}_\beta, \bar{p}_e} \quad -\bar{p}_\beta^T \ln \bar{p}_\beta - \bar{p}_e^T \ln \bar{p}_e$$

$$\begin{aligned} s.t. & \\ & \bar{g}\left(\bar{p}_\beta, \bar{p}_e, \bar{z}_\beta, \bar{z}_e, \bar{y}\right) = \bar{0} \\ & \iota \cdot \bar{p}_\beta^T = 1 \\ & \iota \cdot \bar{p}_e^T = 1 \end{aligned}$$

where the resulting probability vectors $\bar{p}_\beta, \bar{p}_e$ define the production function parameters $\bar{\beta}$ and stochastic sampling errors $\bar{e}$ in terms of expectation (i.e. $E(\bar{\beta}) = \bar{p}_\beta^T \bar{z}_\beta$, $E(\bar{e}) = \bar{p}_e^T \bar{z}_e$). The estimating equations $\bar{g}\left(\bar{p}_\beta, \bar{p}_e, \bar{z}_\beta, \bar{z}_e, \bar{y}\right) = \bar{0}$ are appropriate functions of the data $(\bar{y})$ and user-defined 'support' vectors $(\bar{z}_\beta, \bar{z}_e)$ that

---

[3] We are indebted to Wolfgang Britz for the original idea, and other helpful comments.
[4] The ε also prevents a degenerate dual solution (see Howitt, 1995)

define the space that can contain non-zero probability mass. ι is a unit vector which defines the usual adding up property of the probability measures.

Before the GME reconstruction program can be solved, however, these support values have to be defined for each parameter and error term. To ensure that the set of support values spans the feasible solution set, we define the support values for the production function parameters $(\bar{z}_\beta)$ as the product of a set of five weights and functions of the average Leontieff yield over the data set, for a particular crop/input combination. The support values for the error terms $(\bar{z}_e)$ are defined by positive and negative weights that multiply the right-hand side values of the equation defined above for the expected vector of sampling errors[5].

If the quadratic production function is defined as:

(2) $$y_j = \alpha_j' x_j \; - \; 0.5 \;\; x_j' \;\; Z \;\; x_j$$

then, the resulting GME reconstruction problem becomes:

(3)

$$Max \;\; \Sigma_{i,p} pa_{i,p} \; -\ln pa_{i,p} \; + \; \Sigma_{i,ii,p} pl_{i,ii,p} \; -\ln pl_{i,ii,p} \; + \Sigma_{n,i,p} pe1_{n,i,p} \; -\ln pe1_{n,i,p}$$
$$+ \Sigma_{n,p} pe2_{n,p} \; -\ln pe2_{n,p}$$

$Subject \;\; to:$

$$sh_{n,i} \; = \; \Sigma_p pa_{i,p} \;\; {*}za_{i,p} \; - \; [\; \Sigma_{ii} \;\; \Sigma_p(pl_{i,ii,p} \;\; {*}zl_{i,ii,p}) \; * \; \Sigma_p(pl_{i,ii,p} \;\; {*}zl_{i,ii,p}) {*} x_{n,ii}]$$
$$+ \; \Sigma_p \;\; pe1_{n,i,p} \;\; {*}ze1_{n,i,p}$$

$$yld_n = \; \Sigma_p pa_{i,p} \;\; {*}za_{i,p} \; - \; [\Sigma_{i,p} \;\; \{(\Sigma_{ii,p} \;\; pl_{i,ii,p} \;\; {*}zl_{i,ii,p}) \; * \;\; pl_{i,ii,p} \;\; {*}zl_{i,ii,p} {*} x_{n,ii}\} {*} x_{n,i}]$$
$$+ \; \Sigma_p \;\; pe2_{n,p} \;\; {*}ze2_{n,p}$$

$$\Sigma_n\{ \; \Sigma_p pe1_{n,i,p} \;\; {*}ze1_{n,i,p}\} \; = \; 0$$

$$\Sigma_n\{ \; \Sigma_p pe2_{n,p} \;\; {*}ze2_{n,p}\} \; = \; 0$$

$$\Sigma_p pa_{i,p} \; = \; 1, \; \Sigma_p pl_{i,ii,p} \; = \; 1, \; \Sigma_p pe1_{n,i,p} \; = \; 1, \; \Sigma_p pe2_{n,p} \; = \; 1.$$

Curvature is added by solving for the parameters of the Cholesky decomposition of the quadratic matrix L where Z= LL′, and constraining the diagonal Cholesky parameters to be nonnegative, for details see (Paris & Howitt, 1999). The objective function is the usual sum of the entropy measures for the parameter probabilities for the Cholesky decomposition of the quadratic matrix and the vector of linear terms.

---

5 while these are usually spaced symmetrically about zero, by practitioners of GME, this is not enough to ensure that the model error is zero in expectation. Further comment on this point is made in the paper.

Following the normal GME procedure, the entropy of the error term probabilities is also maximized. The first equations are the first-order conditions that set the cost ratio[6] equal to the marginal physical product. If some inputs are restricted and the PMP calibration stage is used, the input cost in the first order equation will include the resulting shadow values as well as the nominal input price.

The second set of equations fit the production function to the observations on average yield. While it is not normal in econometric models to include average product equations, we think that the information contained in this constraint is particularly important for two reasons. First, information on yields is likely to be the most reliable data the farmer can give. While farmers are often doubtful and reluctant about stating their costs of production to survey enumerators, they always know their yields and hold them as a point of pride in their farming ability. Second, while the marginal conditions are essential for behavioral analysis, policy models also have to accurately fit the total product in order to be convincing to policy makers and to correctly estimate the total impact of policy changes on the environment and the regional economy. Fitting the model to the integral as well as the marginal conditions, therefore, improves the policy precision of the model.

The next two equations are not found in the standard GME specification. Since they require the sum of the errors on each equation sum to zero over the observations, they can be thought of as moment conditions that force the resulting small sample estimates to be unbiased. Even though the support values for the error terms are centered around zero, this does not ensure that the resulting maximum entropy solution is centered around zero, as the relative weight of the error term probabilities in the solution depends on the number of parameters and observations in the reconstruction[7].

The remaining equations in the reconstruction program are the standard adding up constraints on the parameter and error probabilities, as explained previously. Due to the separability assumption on the production functions, if the shadow value of the constraining allocatable resources is included in the input cost, the reconstruction problem can be solved rapidly by looping through individual production functions.

We recognize the sensitivity of our parameter estimates to the choice of supports that we specify for the parameter space. Many practitioners of GME estimation overlook the importance of this issue, and assume a general insensitivity of parameter estimates to the researcher's choice of support specification. Paris & Caputo (2001) have demonstrated the importance of support specification, both analytically, as well as through Monte Carlo studies. While we have not implemented the techniques proposed by others to get around this shortcoming of GME estimation (Paris, 2001; Marsh & Mittlehammer, 2001), we have placed a priority on exploring them in future work. We do, however, pay close attention to an even more serious source of bias in GME estimates, by imposing moment constraints, similar to those used in GMM (Hansen, 1982) or Empirical Likelihood estimation (Qin & Lawless, 1994). We find that this has not been explicitly addressed in the empirical literature employing GME methods, and should be of concern to practitioners.

[6] Defined as the ratio of input cost over output price

[7] See additional development of this point in Howitt & Msangi (2002)

Calculating Comparative Static Parameters for the Model

The quadratic production function model has convenient properties for calculating policy parameters. Note that the Hessian resulting from the constrained profit maximization problem (1), is simply: $\frac{\delta^2 \Pi}{\delta x_{ij}^2} = -Z$, which will be useful in the following derivations.

Calculating the Derived Demands for Inputs

For simplicity, we will use the unconstrained profit term for a single crop.

(4)

$$\Pi = p \ (\alpha' x - 0.5 \ x' \ Z \ x) \ -\omega' x$$

$$\frac{\delta \Pi}{\delta x_i} = p(\alpha - Z \ x) - \omega = 0$$

$$Z \ x_i = \alpha - \frac{\omega_i}{p}$$

$$x_i^* = Z^{-1}\alpha - Z^{-1}\frac{\omega_i}{p}$$

$$\textit{Define the demand slope } g_i = -Z^{-1}\frac{1}{p}$$

$$x_i^* = a_i + g_i \ \omega_i$$

From the above equations, it is clear that if we can invert the Hessian of the maximized profit expression (- Z^{-1}) we can calculate the derived demand for each input for each crop as a linear function of input and output price.

The elasticity of the input demand follows directly from :

(5) $$\textit{The elasticity is } \eta_i = g_i \ \frac{\omega_i}{x_i^*}$$

This elasticity is based on a single crop. For the usual multi-output case we weight the individual crop contribution by their relative resource use to arrive at a weighted elasticity for the resource.

Calculating Supply Functions and Elasticities

Since production is a function of optimal input allocation and we now have the input demands as a function of input and output price, we can derive the output supply function by substituting the optimized input derived demands into the production function and simplify in terms of the output price. Going back to the derived demand and production function formulae:

(6)

$$x_i^* = Z^{-1}\alpha - Z^{-1}\frac{\omega_i}{p}$$

and

$$y^* = \alpha' x^* - 0.5 \quad x^{*'} \quad Z \quad x^*$$

$$Defining \;\; r_i = \frac{\omega_i}{p}$$

$$Using\; the\; vectors\; \alpha \; and \; r$$

(7)

$$y^* = \alpha' \; Z^{-1}(\alpha - r) - 0.5 \; (\alpha - r)' \; Z^{-1} Z \; Z^{-1}(\alpha - r)$$

multiplying out and collecting α *and* r *terms yields*

$$y^* = 0.5 \;\; \alpha' \;\; Z^{-1}\alpha - 0.5 \;\; r' \;\; Z^{-1} r$$

$$y^* = \hat{\alpha} - 0.5 \left(\frac{\omega_1}{p} \; ... \frac{\omega_I}{p} \right)' Z^{-1} \left(\frac{\omega_1}{p} \; ... \frac{\omega_I}{p} \right)$$

(8)

$$y^* = \hat{\alpha} - \frac{1}{2p^2}\omega' Z^{-1}\omega$$

$$\frac{\delta y^*}{\delta p} = \frac{1}{p^3}\omega' Z^{-1}\omega$$

$$\eta_S = \frac{\delta y}{\delta p}\frac{p}{y} \quad = \frac{1}{p^3}\omega' Z^{-1}\omega \frac{p}{y}$$

$$\eta_S = \frac{1}{p^2 y}\omega' Z^{-1}\omega$$

Calculating Elasticities of Input Substitution

There are many elasticities of substitution with different advantages and disadvantages. To demonstrate that we can obtain crop and input specific elasticities of substitution we use the classic Hicks elasticity of substitution defined by Chambers (1988, p. 31) as:

(9)

$$y^* = f(x_1 \; .... \; x_I)$$

$$\sigma_{1,2} = \frac{-f_1 \quad f_2(x_1 f_1 \;\; + x_2 f_2)}{x_1 \quad x_2 \quad (f_{11} f_2^2 - \; 2 f_{12} f_1 f_2 + f_{22} f_1^2)}$$

where f_i *and* f_{ij} *are derivatives*

The above derivations show that if the production function is quadratic and the Hessian is invertible, then it is possible to calculate all the standard econometric comparative static qualitative production measures that apply to the calibrating data set and reflect the average shadow values for restricted or incompletely priced inputs.

We note that the supply functions, derived input demands, their associated elasticities, and the elasticities of substitution are obtainable from a data set of any size from one observation upwards. Clearly the reliance on the support space values and the micro theory structural assumptions is much greater for minimal data sets. However the approach does enable a formal approach to disaggregation of production estimates, since the specification of the problem is identical for all sizes of data sets.

So far, this analysis has used the quadratic production function. Two other functional forms that are widely used are the Generalized Leontieff and Translog production functions. As would be expected, the ME empirical reconstruction methods outlined in this chapter apply equally well to these production functions, provided the input data for the trans-log function is bounded for a local optima.

IV. The Empirical Reconstruction of Regional Crop Production in California.

The empirical setting in which we will present our reconstruction approach is that of the California Statewide Water and Agricultural Production (SWAP) model (Howitt et al., 2001). In this paper, however, we will go beyond the deterministic Maximum Entropy reconstruction in SWAP (based on a single year of statewide data), and employ GME over several years of data from a subset of the original SWAP regions.

SWAP is a multi-input, multi-output economic optimization model that is disaggregated into 24 regions that span the main agricultural regions of California. This level of disaggregation is based on the way that agricultural data is collected in California, and how water allocation institutions and agencies vary by regions.

The SWAP model specifies agricultural water demands on a monthly basis and extrapolates the model to consider agricultural water demands for the year 2020. Irrigated acreage, by crop and region, for the base year 1995 and forecasts for the year 2020 are based on California water agency data. This specification allows for a robust representation of alternative water management policies that can be interpreted on a statewide basis.

Regional crop prices, yields and areas grown were based on annual county agricultural commissioner reports. The data used in our reconstructed GME-based model includes the years 1993 to 1998, with prices normalized to 1992 levels.

Data Restrictions

Ideally, production models are reconstructed from a consistent time series of regional data, which includes all the crop inputs and outputs and their associated prices. Unfortunately, such rich, consistent data sets are rarely available. In some cases, comprehensive cross-section survey data is available, but it is rarely collected for more than one year. Given these restrictions, we have to use data collected annually by regional public agencies.

The data available is similar to many disaggregated production data sets collected consistently by public agencies. Typically, such data sets contain data that is available on a regional basis for crop yields, prices, and acres harvested, but rarely is available for other input use on a crop and regional basis. In this example that focuses on irrigation water use and its derived demand we have to generate estimates of the crop water allocation and associated capital inputs. These are derived by combining the crop acreage data and total surface water allocated to regions with crop water requirements based on the regional climate, and the efficiency of water application in the region. Other authors (Lence & Miller, 1998b) have found the Entropy principle to be useful in recovering activity-specific input usage from aggregate data.

The production functions in this paper are reconstructed from four years of data (1995 – 1998). However, we are faced with data inconsistencies that are commonly encountered when using empirical farm data for estimation. Specifically, the problem is that the observed crop allocations, based on the raw data, sometimes violate the assumption of efficiency and optimal decision-making. Since SWAP, like the majority of economic models, is predicated on the assumption of efficient decisions and bounded rationality, the empirical data has to be reconciled with economic efficiency before we can proceed with reconstruction of the parameters. The empirical task is to calibrate our model to account for all the crops that are observed to be grown, even those that, based on the raw data, appear to lose money. In short, we have to adjust the raw data to make it consistent with informed and rational allocations.

The most logical explanation for growing crops that appear to lose money is that there is some added benefit from the crop to the producers that is not reflected in the observed data. This increased profitability can take the form of either reduced costs for that crop or increased returns in other crops. The marginal opportunity cost of production for apparently “inefficient” crops that require labor and machinery at a different time than the main "cash" crops will be below the input opportunity costs for the main "cash" crops. Crops with this characteristic are colloquially termed a "filler" crops by some farmers. A more common reason for growing apparently “inefficient” crops is the rotational benefits on crop yields that many low value "rotational" crops confer on cash crops in subsequent years. Technically this effect is an agronomic complementarity that adds to long-term revenue, but is not reflected in the annual crop costs and returns. In both econometric and programming models, these inter-temporal effects are usually addressed by arbitrary data adjustments to the cost or returns of the relevant crops. An alternative, and equally unsatisfactory method, is to impose fixed rotational proportions as constraints among the crops. These ad hoc adjustments do not use a consistent definition of inefficiency, or use a specified measure of adjustment across all crops and inputs.

The capital costs used to calibrate SWAP are restricted to those used in irrigation since this is the particular focus of the model. The annual variable cost of capital in production therefore represents the annual irrigation system cost per acre, and is a combination of labor, management, capital costs and an associated irrigation technology that yields a given irrigation efficiency. The variable capital cost defines a functional relationship between the cost of irrigation technology and improvements in water use efficiency. Thus, investments in “better” irrigation technology result in an increase in irrigation efficiency that uses less applied water to achieve the same yield. In the SWAP model, prior estimates of CES isoquant functions (USBR & Hatchett, 1997) are used to calculate the variable capital cost that is implied by a given physical efficiency of water use for a specific region and crop. The base year irrigation efficiency is calculated from the ratio of the regional ET divided by the observed applied water. This efficiency value is used in the CES function to solve for the appropriate capital cost using parameter estimates presented in CVPM.

The cost per acre-foot of water for each region uses a weighted average cost of the aggregate supply of water for all districts within the region. Solving the production optimization problem generates the annual amount of applied water that is allocated by crop and region. The annual quantity is apportioned by month, based on monthly crop water requirements, which have been identified for each crop and region using the Department of Water Resources Consumptive Use Model (1997).

Production Function Specification

In the original SWAP specification, each region has a different production function for each of the crops produced. Within a region the production of different crops is connected by the restrictions on the total land and water inputs available and the land cost function. The wide range of agricultural production inputs has been aggregated and simplified to just land, water, and capital.

The production function is written, in general, as:

$$(10) \quad y = f(x_1, x_2, x_3)$$

The specific quadratic form used in the SWAP model has the form:

$$(11) \quad y = [\alpha_1, \alpha_2, \alpha_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - [x_1 \quad x_2 \quad x_3] \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

where y is the total regional output of a given crop and x_i is the quantity of land, water or capital allocated to regional crop production. The quadratic matrix zeta (Z) captures decreasing marginal productivity of inputs, as well as interaction effects between inputs. Second-order conditions for the production problem require that the zeta matrix is positive-definite, and are implemented by imposing necessary conditions on the Cholesky decomposition of the zeta matrix.

The land allocated to different crops is subject to substitution and complementary relationships between crops. These effects are due to the inter-dependence of crop production rotations, the heterogeneity of land and its restricted quantity for many farm enterprises

The full problem defined over G regions and i crops in each region for a single year is :

$$(12) \quad \begin{aligned} &\text{Max } \Sigma_G \ \Sigma_i \ p_i \ f_{Gi}(x_1, x_2, x_3) - \omega_1 x_1 - \omega_2 x_2 - \omega_3 x_3 \\ &\text{subject to } \quad \Sigma_{Gi} \ x_{1Gi} \le X_1 \ (\text{Land}) \\ &\qquad\qquad\quad \Sigma_{Gi} \ x_{2Gi} \le X_2 \ (\text{Water}) \end{aligned}$$

where the total annual quantities of irrigated land and water (X_1 and X_2) are limited in each region and must be optimally allocated across crops grown in that region. By changing the RHS quantity of water available on the constraint, we generated a derived demand function for each region[8], which were then used to define agricultural demand nodes for water within the economic –engineering model (CALVIN[9]). Since CALVIN requires that water is valued on a monthly basis, the model specification was modified to give monthly valuations of water, by specifying monthly crop water requirements, as a proportion of the total annual consumptive use applied in each month. For the purposes of this paper, we have kept the optimization on the basis annual resource requirements and allocations.

Reconstruction of the Economic Model

Calibration of the full set of parameters for the production function with three inputs requires that each regional crop be parameterized in terms of nine parameters, three for the linear terms, and six for the quadratic matrix. After imposing the required symmetry restrictions for the off-diagonal terms, in a single set of base year data the number of equations are limited to three first-order conditions, which represent the underlying behavioral assumption of optimizing behavior with respect to the three production inputs. Given the small sample of four years, we have twelve observations on input allocations, and three observations on output, for a total of 16 observations. The resulting disaggregated production function has 7 degrees of freedom.

In addition, given the methods used to generate the water application and irrigation capital input levels, there is likely to be considerable collinearity between the inputs for a given crop and region. Fortunately, the GME approach that we use for the reconstruction is robust under collinearity and low degrees of freedom.

The SWAP model is reconstructed in three stages. In the first stage, a linear programming (LP) model is constructed for each region that incorporates all the available data on cropping acreages, annual water use, yields, output prices and input costs and quantities. The LP model is maximized subject to land and water constraints and also a set of constraints that calibrate the model to the observed land use and production quantities in each region. This initial stage is the same as that used in the positive programming approach (Howitt 1995). Stage two consists of reconstructing the production function by Maximum Entropy, using the calibration constraint shadow values to define its curvature; which is followed by the solution of the non-linear constrained maximization problem (12) in the third stage.

Where land is limiting and an adequate measure of the rental rate for land is not available, the opportunity cost of land must be inferred from the shadow value of the crops grown. However, for multicrop systems where rotational interdependencies among crops change their marginal contribution to rental returns, the crops need to be divided into those that are net users of attributes (weed and disease control, or fertility) from other crops and those that are net contributors to the farm productivity. The third group is assumed to neither contribute to, nor reduce farm productivity. We can term these three groups of crops as cash crops, rotational crops, and filler crops.

[8] note that these can be obtained directly from the inverted Hessian, as explained previously
[9] California Value Integrated Network (http://cee.engr.ucdavis.edu/faculty/lund/CALVIN/)

Taking the example of Imperial Valley, from our SWAP data set, we can make a comparison between the ranking of each crop category according to the net income earned per acre, as well as to the shadow values derived from the 1st stage LP problem (13)

$$\max_x \; [p' - c']x$$
$$s.t.$$
$$Ax \leq b$$
$$x \leq \bar{x}(1+\varepsilon)$$
$$x \geq \bar{x}(1-\varepsilon)$$

where $\bar{x}$ is the vector of observed crop acreages, p' and c' are the revenues and costs per acre for each crop, A is the usual Leontief input-output matrix, b is the vector of available inputs and ε is a small perturbation. The vector of shadow values λ^U and λ^L are associated with the upper and lower bound constraints on observed acreage, respectively, and are shown below, alongside the cash net income generated per acre by each crop.

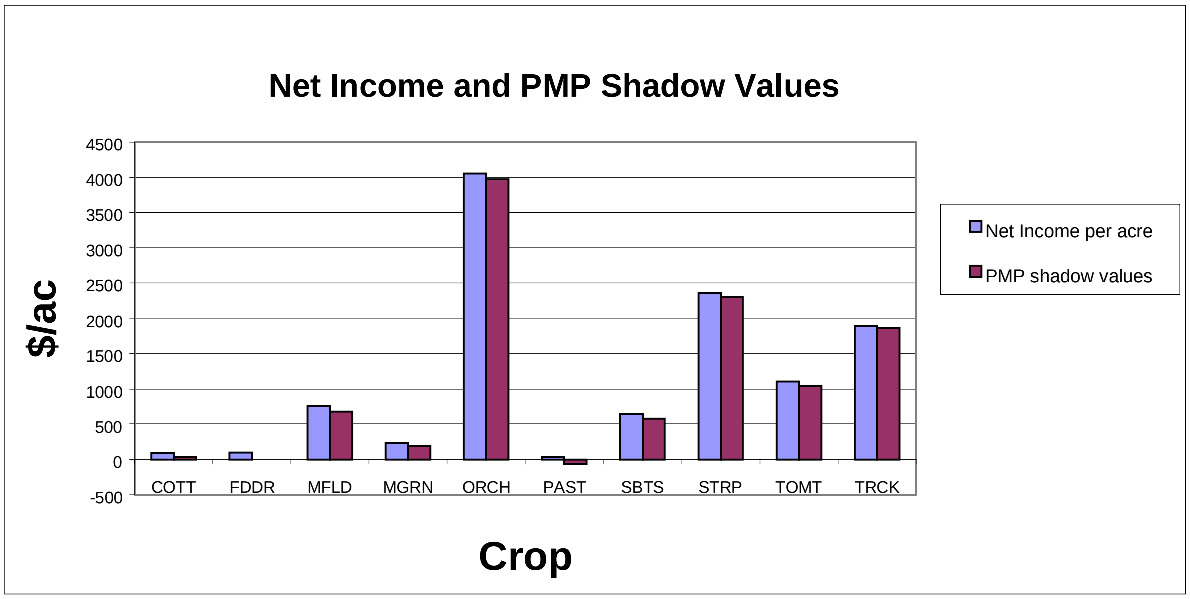


Figure 1.

The rankings are, of course, similar, given that the shadow values are defined in terms of the objective function value. Pasture gives a negative shadow value on its calibration constraint, while Fodder has none (it, instead determines the opportunity cost of land).

A similar ranking, however, can also be derived from applying non-parametric data envelopment analysis (DEA) to our data. The non-parametric mathematical programming approach to measuring efficiency in production was pioneered by Farrell (1957), and has been the basis of a vast literature on productive efficiency of firms, and other productive decision-making units (*Färe* et al., 1985, 1994; Charnes et al., 1978). Farrell's concept of allocative efficiency of productive inputs among firms within an industry ties in very well with the Pareto-Koopmans concept of efficiency under decentralized allocation (Sengupta, 1989), and we can represent the DEA approach with the following LP problem, following the formulation in Sengupta:

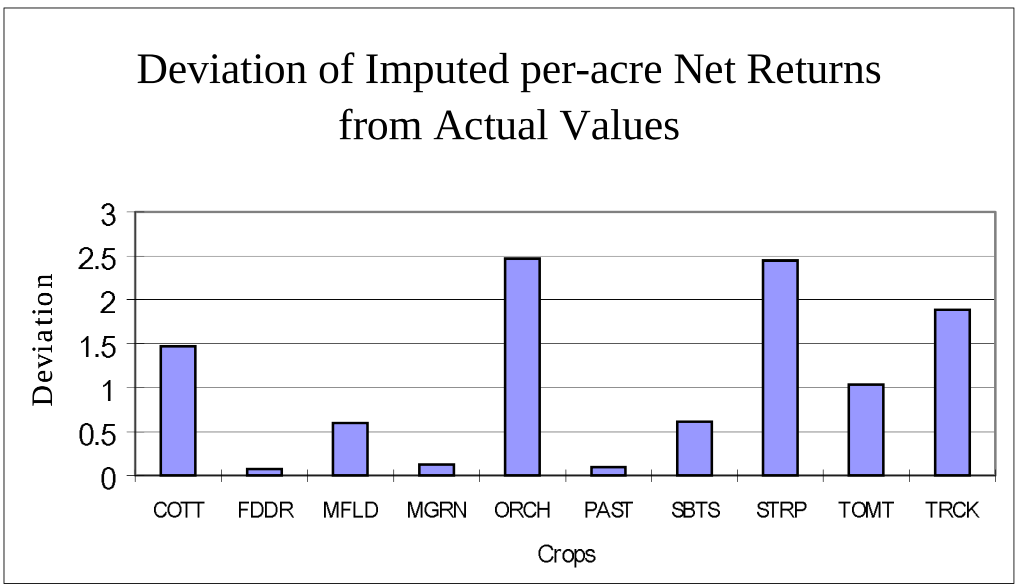
(14)

$$\min_{\alpha,\beta} \sum_i \beta_i \bar{x}_{ik}$$
$$s.t. \;\; \alpha y_k = 1$$
$$\sum \beta_i \bar{x}_{ij} \geq \alpha y_j$$

where (i,j) are respective indices for the set of inputs and output activities, and k is a particular output (crop). The 'reference' activity, k, serves to define the numeraire condition for a Walrasian equilibrium, which assigns prices to the outputs and inputs of productive firms within an industry. In our formulation of this problem, we treat the crop enterprises that we observe in our data set as decision-making firms, and repeatedly solve the LP problem above, using each crop, in turn, as the reference output j', so as to obtain a set of imputed Walrasian "efficiency" prices for each crop and input $(\hat{\alpha}_{j'}, \hat{\beta}_{ij'})$. From these imputed prices, we can then calculate the imputed per-acre net return for each crop activity as: $nr_{imp} = \hat{\alpha}_j \cdot yld_j - \sum_i \hat{\beta}_{ij} \cdot a_{ij}$, where yld_j is the crop yield and a_{ij} is the Leontieff coefficient corresponding to input i and output j. This can then be compared to the true per-acre net return of each crop activity, as calculated previously for the PMP 1st stage LP program as: $nr_{true} = p_j \cdot yld_j - \sum_i c_{ij} \cdot a_{ij}$, where p_j and c_{ij} are, respectively, the true output prices and input costs observed in the data.

By taking the absolute value of the difference between these calculated net per-acre returns[10], we're able to obtain the ranking shown below:

Figure 2.



[10] and scaling down the actual net-returns calculated in the data by an appropriate scale, so that they're comparable in magnitude

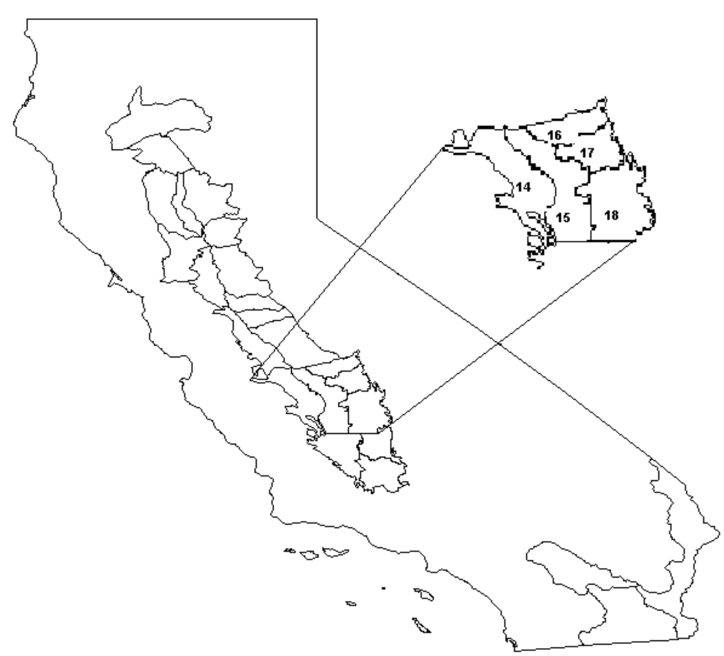
Which corresponds quite closely with the ranking given previously by the upper and lower bound PMP shadow values, although cotton is somewhat higher in prominence. This shows that the PMP-derived measure of 'revealed' efficiency in the allocation of land between cropping activities, corresponds closely with measures derived from traditional non-parametric measures of allocative efficiency in production and their deviation with net returns to land we observe in the data. This suggests to us that PMP constitutes a valid basis for characterizing land allocation behavior in production, and can be further augmented by the GME reconstruction framework presented here.

Reconstructing a Five-Region Production Model

Data from five regions in California's central San Joaquin valley are used to illustrate disaggregated reconstruction process. The reconstruction sequence proceeds by first solving the constrained calibration in equation (1) and using the resulting shadow values to define the LHS of the first-order conditions for input allocation in equation (3), and solving the generalized maximum entropy problem. The resulting probabilities from the GME solution are used to generate the expected values of the production function parameters α_i and z_{ij} . The regional production problem defined in equation (12) is solved using the production function specified in equation (11).

The five contiguous regions shown in figure(3) range from 1,022 to 51 thousand irrigated acres and have different institutions, water prices , and drainage conditions. The reconstruction problem was defined as a series of linked nonlinear optimization problems and solved by GAMS (Brooke et al., 1988), taking 90 seconds on a 0.75 megahertz PC. The problem requires the reconstruction of 33 three-input quadratic production functions, one for each regional crop grown.

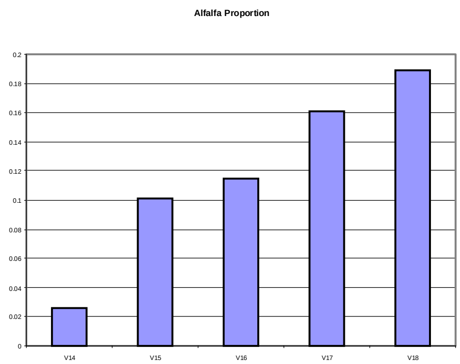
Figure 3



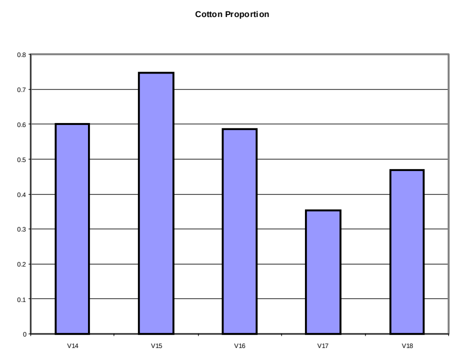
Despite being contiguous regions in the same valley, there are considerable differences in cropping patterns. Figures 4 (a) – (d) show the regional differences for four of the principle crops.

Figures 4a and 4b

Alfalfa

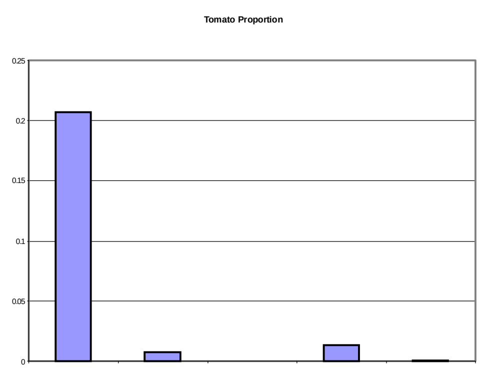


Cotton

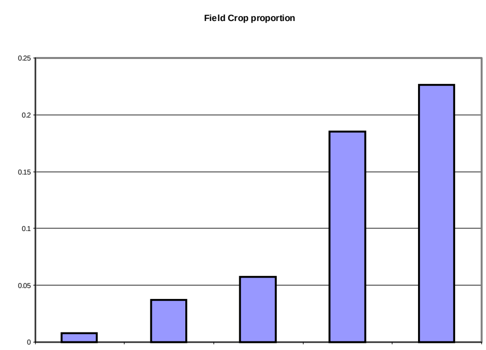


Figures 4c and 4d

Tomato



Field Crops



The five regions despite their proximity, show a wide variation in crop selection due to differences in soil-type, microclimate, water quality, and water constraints.

The goodness-of-fit of the simulated GME regional models can be measured by a measure equivalent to the familiar R^2 parameter[11] applied to the results from

[11] Defining $\hat{y}_i, y_i, and\ \bar{y}$ respectively as the simulated production, actual production, and mean production, the simulation R^2 = $1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2}$

simulations based on the reconstructed crop production functions. Two key measures are the levels of regional crop production and the associated regional cropland allocations. The results show that the model simulations explain a large proportion of the deviation from the mean level of crop production, and a lower, but acceptable proportion of the land allocation. While the production R^2s are comparable in many crops with large sample econometric results, it should be borne in mind that we are applying large sample criteria to reconstructions (estimates) based on very small collinear data sets.

When the R^2 parameters are weighted by the regional crop proportion to the total cropped area they show the overall fit of the simulated model. The weighted production R^2 parameter for the disaggregated model has a value of 0.53, while the simulated land area by crop and region has a weighted value of 0.21. These aggregate parameters indicate the amount of additional information in the simulation model over the disaggregated sample mean.

In absolute terms the disaggregated simulation model captures the regional input allocation well, with a weighted percentage error of 11.95% for land, 17.24% for water, and 17.36 % for capital. The average absolute percentage error of crop yield prediction is 8.86%

Table 1.
Disaggregate Model R^2 Values for Regional Production

	Cotton	Alfalfa	Field	Grain	Tomato	S-Beets	Truck
Regions							
V14	0.892	-2.291	0.181	-0.537	-0.287	1.000	0.808
V15	0.878	-2.157	0.997	0.331	0.864	0.993	0.527
V16	0.869	-9.147	-1.686	0.756			0.710
V17	0.921	0.559	0.989	-3.595	0.157	0.998	0.429
V18	0.919	0.794	0.996	0.355	0.946	1.000	0.937

The simulation R^2 measures show that, for most regions and crops, the simulated production has good explanatory power and fit. What is surprising, and currently inexplicable, is how the regional disaggregated model can fit a crop well for four regions and be completely wrong for the fifth region. See, for example, the case of field crops in table 1. The explanation for this may be that the shared simple price and yield expectation scheme are inappropriate for that region. This problem needs further investigation. The R^2 for simulated land allocation shown below in table 2 indicates less precision in fit than that of the production results. This is to be expected, as the simulated allocation of land input is an additional stage removed from the estimation than the production levels, with the consequent increase in potential error.

Table 2.
Disaggregate Model R^2 Values for Regional Land Allocation

	Cotton	Alfalfa	Field	Grain	Tomato	S-Beets	Truck
Regions							
V14	0.797	0.193	-1.380	-0.126	-0.532	-0.135	-0.628
V15	0.834	-1.933	-0.476	-0.537	0.467	-1.120	-2.145
V16	0.814	0.363	0.752	0.172			-0.710
V17	0.350	-0.750	0.246	0.268	-1.622	-0.351	-0.937
V18	0.509	-1.183	-0.355	-0.989	0.200	-0.374	0.393

Being that the primary use of this regional model is to estimate the response of farmers to changes in water price or availability, it follows that the elasticity of demand for water is the most important policy parameter derived from this study. If the elasticity differs over regions there will be gains in policy precision for the disaggregated regional model.

Table 3.
Regional Input Demand Elasticities.

	LAND	WATER	CAPITAL
V14	1.502	0.841	0.583
V15	1.614	0.628	0.392
V16	1.318	0.537	0.364
V17	1.296	0.429	0.377
V18	1.405	0.501	0.425

Table 3 shows that all the water derived demands are all within the expected inelastic range, but the regional difference in elasticity between the highest and lowest is almost two-fold. An average elasticity over the regions would distort the response to water policy, given the substantial differences in the size of the regions and their elasticities. The two other inputs show a smaller, but significant range of elasticities, further justifying the use of the disaggregated model.

Table 4.
Regional Output Supply Elasticities

	Cotton	Alfalfa	Field	Grain	Tomato	S-Beets	Truck
Regions							
V14	1.079	0.683	0.107	0.308	2.013	0.083	0.317
V15	1.293	0.141	0.148	0.363	2.514	0.082	0.411
V16	1.109	0.723	0.117	0.299			0.114
V17	1.258	0.643	0.120	0.352	1.688	0.127	0.282
V18	1.346	0.545	0.113	0.398	1.577	0.068	0.880

The supply elasticities shown in table 4 are within the expected range, with the key crops of cotton and tomatoes being moderately elastic. The very inelastic supply response of sugar beets may be due to the current restriction on sugar beet processing capacity in the region. Some crops, such as cotton have a fairly uniform elasticity across regions indicating there's little qualitative advantage to having a disaggregated model. However, other crops such as alfalfa and truck crops show upwards of a five-fold difference in elasticity values, indicating that a significant benefit could be gained from a disaggregated analysis.

Table 5 shows the Hicks elasticities of substitution between land water and irrigation capital for a specific region (region V15). Except for tomatoes, all the substitutions are in the expected (relatively inelastic) range. Since the common requirement to calculate elasticities of substitution is the ability to reconstruct the Hessian, other more sophisticated substitution measures can also be calculated.

Table 5.
Regional Hicks Elasticities of Substitution

	WATER	CAPITAL
Cotton.LAND	0.536	0.306
Cotton.WATER		0.671
Alfalfa.LAND	0.609	0.533
Alfalfa.WATER		0.459
Field.LAND	0.578	0.641
Field.WATER		0.547
Grain.LAND	0.303	0.219
Grain.WATER		0.385
Tomato.LAND	1.597	1.260
Tomato.WATER		1.160
S-Beets.LAND	0.346	0.529
S-Beets.WATER		0.546
Truck.LAND	0.808	0.923
Truck.WATER		0.763

Reconstruction of an Aggregate Production Model

The performance of an aggregate model over all the five regions and four time periods is tested by reconstructing an aggregate GME production function for each crop. A regional dummy variable is also included to allow for a linear shift in production between the differently-sized regions. The aggregate specification is identical to the regional form, but with the added benefit of the linear dummy variable. The number of parameters for each crop production function is now 10, and the number of observations used for each estimation is 80 giving the resulting reconstruction greater degrees-of-freedom. Despite the larger sample size and the regional dummy variables, the ability to predict regional crop production or land-use allocation is abysmal. Tables 6 and 7 show the equivalent measures of fit (R^2) for the aggregate model, as tables 1 and 2 did for the disaggregated model.

Table 6.
Aggregate Model R^2 Values for Regional Production

	Cotton	Alfalfa	Field	Grain	Tomato	S-Beets	Truck
Regions							
V14	0.677	-2573.590	-1022.030	-8.222	-377.648	0.466	-69.145
V15	-0.022	0.096	-2.697	-0.412	-708.790	0.725	-18.261
V16	-9.907	-1185.980	0.865	-201.554			-17.689
V17	-1.698	-0.955		-273.980	-1143390.000	-7278.790	-37.903
V18	-2.904	-19.837	-12.723	-0.140	-94245.800	-17.856	-7583.600

Table 7.
Aggregate Model R^2 Values for Regional Land Use

	Cotton	Alfalfa	Field	Grain	Tomato	S-Beets	Truck
Regions							
V14	-0.746	-2055.250	-3618.360	-3.765	-339.791	-0.081	-78.336
V15	0.154	-4.533	-47.123	-1.552	-692.588	-1.055	-32.596
V16	-3.909	-106.640	-0.532	-28.513			-99.504
V17	-0.838	-1.018		-24.178	-41206.000	-4112.380	-112.878
V18	-0.691	-24.982	-14.699	-28.914	-69459.600	-25.483	-8982.340

In reviewing tables 6 and 7, remember that any value that is less than one indicates that using the disaggregate average data value will yield a better disaggregate fit than the simulated output of the aggregate model. For the production levels, only 18% of the regional values meet this criteria. For land allocation, the results are even worse with only one of the regional land allocations beating the average value.

V. Conclusions

This paper shows that, by using a combined PMP and GME approach, it is possible to reconstruct flexible form production function models from a data set of modest size. A researcher can reconstruct a similar theoretically-consistent flexible form production model using a data sets that range from “LP budget data” to full econometric data sets with standard degrees of freedom. The convergence of GME estimates to conventional estimates as the sample size increases means that as the data set is expanded there is a continuum between optimization and econometric models.

The reconstructed production models yield all the comparative static properties and parameters of large sample models, thus enabling the input demands, output supplies and elasticities of substitution to be calculated directly instead of by the parametric methods traditionally used by programming models. The effect of any constraints on production is directly incorporated into the estimates, through the addition of the calibration constraint shadow values to the nominal prices of the allocatable resources.

Due to the relative sparseness of the data set used in this paper (only four years), the fit of the resulting production model significantly depends on the definition and range of prior support values for the parameters, and the expectation structure assumed for the farmer’s expected yields and prices. Nonetheless, we feel that this use of prior information on the farming system and process of production is a valuable source of modeling information that should be formally included. An advantage of modeling production functions (over dual cost functions) is that the variables (being quantities of input and output) can be readily understood and used by researchers in other disciplines[12].

In this example the aggregation bias in the aggregate model swamped any gains in reducing the small sample bias. The disaggregated model yielded greater precision and regional response. The gain from disaggregation of production models is an

[12] Such as the civil engineers with whom we collaborated with on the CALVIN engineering-economic modeling project.

empirical result that needs substantially more testing before one can conclude that it is a common phenomenon, however in this California cropping example the empirical results are conclusive.

References

Alig, R.J., D.M. Adams and B.A. McCarl. " Impacts of Incorporating Land Exchanges Between Forestry and Agriculture in Sector Models." *Journal of Agricultural and Applied Economics*, 30:2 (1998): 389-401.

Antle, J.M. & S. Capalbo. "Econometric-Process Models of Production." *American Journal of Agricultural Economics,* 83 (May 2001): 389-401.

Brooke, A., D. Kendrick, & A. Meeraus. "GAMS: A Users's Guide". The Scientific Press, 1988.

Charnes, A., W.W. Cooper & E. Rhodes. "Measuring the Efficiency of Decision-Making Units." *European Journal of Operational Research*, 2 (1985): 429-444.

Chambers, R. "Applied Production Analysis: The Dual Approach", Cambridge University Press, Cambridge, 1988.

Coelli, T.J. " Recent Developments in Frontier Modelling and Efficiency Measurement." *Australian Journal of Agricultural Economics,* 39:3 (Dec 1995): 219-245.

Desmedt, P. " Maximum Entropy and Bayesian Image reconstruction: Overview of Algorithms" LEM Technical report DG91-14, University of Ghent, 1991.

Färe , R., S. Grosskopf & C.A.K. Lovell. "Production Frontiers." Cambridge University Press, Cambridge, 1994.

Färe , R., S. Grosskopf & C.A.K. Lovell. "The Measurement of Efficiency of Production." Kluwer Academic Publishers, Hingham, MA, 1985.

Farrell, M.J. "The Measurement of Productive Efficiency." *Journal of Royal Statistical Society*, Series A, 120 (1957): 253-290.

Golan, A., G. Judge, D. Miller. "Maximum Entropy Econometrics: Robust Estimation with Limited Data", John Wiley & Sons, Chichester, England, 1996a.

Golan, A., G. Judge & J. Perloff. "Estimating the Size Distribution of Firms Using Government Summary Statistics." *J. Indust. Econ.*, 44 (Mar 1996b):69-80.

Golan, A., G. Judge, S. Robinson."Recovering Information from Incomplete or Partial Multi-sectoral Economic Data." *Rev. Econ. Statistics,* 76 (Aug 1994): 541-49.

Hansen, L.P. "Large Sample Properties of Generalized Method of Moments Estimation." *Econometrica*, 50 (Jul 1982): 1029-1054.

Heckelei, T. & W. Britz. "Concept and Explorative Application of an EU-wide, Regional Agricultural Sector Model (CAPRI-Project)". Presented at the 65th EAAE-Seminar, Bonn. March, 2000.

Howitt, R. E. "Positive Mathematical Programming." *American Journal of Agricultural Economics*, 77(Dec 1995):329-342.

Howitt, R.E.., K.B. Ward, S. Msangi. "Statewide Water and Agricultural Production Model". Written as Appendix N to the report for the California Value Integrated Network (CALVIN) model., 2001. (see: http://cee.engr.ucdavis.edu/faculty/lund/CALVIN/).

Howitt, R.E. & S. Msangi. "Consistency of GME Estimates Through Moment Constraints." *forthcoming Working Paper*, Department of Agricultural and Resource Economics, University of California at Davis, 2002.

Just,R.E., D Zilberman, and E. Hochman "Estimation of Multicrop Production Functions" *American Journal of Agricultural Economics*, 65 (1983): 770-780.

Just, R.E., J.M. Antle.(1990) "Interactions Between Agricultural and Environmental Policies - A Conceptual Framework". *American Economic Review*, 80:2(May):197-202.

Lence, S. H. & D.J. Miller. "Estimation of Multi-Output Production Functions with Incomplete Data: A Generalized Maximum Entropy Approach." *Eur. Rev. Agr. Econ.*, 25(Dec 1998a):188-209.

Lence, S. H. & D.J. Miller. "Recovering Output-Specific Inputs from Aggregate Input Data: A Generalized Cross-Entropy Approach." *American Journal of Agricultural Economics*, 80(Nov 1998b):852-867.

Love, H.A. "Conflicts Between Theory and Practice in Production Economics." *American Journal of Agricultural Economics*, 81 (Aug 1999): 696–702.

Marsh, M.L. & R. Mittlehammer. "Adaptive Truncated Estimation Applied to Maximum Entropy." *Presented at the Western Agr. Econ. Assoc. Annual Meetings*, Logan, Utah, (July 2001) 30 p.

McCarl, B.A., D.M. Adams, R.J. Alig & J.T. Chmelik, "Analysis of Biomass Fuelled Electrical Power plants: Implications in the Agricultural and Forestry Sectors." *Annals of Operations Research*, 94(2000): 37-55.

Mendelsohn, R., W.D. Nordhaus, D. Shaw. “The Impact of Global Warming on Agriculture: A Ricardian Analysis.” *American Economic Review*, 84 (Sep 1994): 753-771.

Mittelhammer. R.C., G.G. Judge, D. J. Miller. “Econometric Foundations” New York : Cambridge University Press, 2000.

Paris, Q. & R.E. Howitt. "Analysis of Ill-Posed Production Problems Using Maximum Entropy." *American Journal of Agricultural Economics*, 80 (Aug 1988):124-138.

Paris, Q. “Multicollinearity and Maximum Entropy Estimators.” *Economics Bulletins*, 3 (no. 11, 2001):1-9.

Paris, Q. & M. Caputo. “Sensitivity of the GME Estimates to Support Bounds.” *Working Paper*, Department of Agricultural and Resource Economics, University of California at Davis, (Aug 2001) 18 p.

Qin, J. & J. Lawless. “Empirical Likelihood and General Estimating Equations.” *The Annals of Statistics*, 22 (1994): 300-325.

Sengupta, J.K. “ Efficiency Analysis by Production Frontiers: The Non-Parametric Approach.” Kluwer Academic Publishers, Hingham, MA, 1989.

US Department of the Interior, Bureau of Reclamation.(1997) “ Central Valley Project Improvement Act: Draft Programmatic Environmental Impact Statement”, Sacramento, California.. *(S. Hatchett main author of CVPM model description in Technical Appendix)*

Wu, J.J. & B. Babcock. “Meta-modeling Potential Nitrate Water Pollution in the Central United States.” *Journal of Environmental Quality*, 28(Nov-Dec, 1999):1916-28.

Zhang, X. & S. Fan. “Crop-Specific Production Technologies in Chinese Agriculture.” *American Journal of Agricultural Economics*, 83 (May 2001): 378-388.