ESTIMATION AND INFERENCE UNDER ECONOMIC RESTRICTIONS

CHRISTOPHER F. PARMETER, KAI SUN, DANIEL J. HENDERSON, AND SUBAL C. KUMBHAKAR

ABSTRACT. Estimation of economic relationships often requires imposition of constraints such as positivity or monotonicity on each observation. Methods to impose such constraints, however, vary depending upon the estimation technique employed. We describe a general methodology to impose (observation-specific) smoothness constraints for the class of linear regression estimators using a method known as constraint weighted bootstrapping. While this method has received attention in the nonparametric regression literature, we show how it can be applied for both parametric and nonparametric estimators. A benefit of this method is that imposing numerous constraints simultaneously can be performed seamlessly. We apply this method to Norwegian dairy farm data to estimate both unconstrained and constrained parametric and nonparametric models.

JEL Classification: C12 (Hypothesis testing), C13 (Estimation), C14 (Semiparametric and non-parametric methods).

1. INTRODUCTION

"One of the most vexing problems applied economists have encountered in estimating flexible functional forms in the production or consumer context is that the theoretical curvature conditions that are implied by economic theory are frequently not satisfied ..." Diewert & Wales (1987, p. 43). So begins one of the seminal papers on imposing curvature conditions in applied econometric settings. Nearly all empirical studies estimate models which face, to some degree, smoothness restrictions provided by economic theory. These theoretical models often lay down numerous constraints that must be satisfied by the functional relationship of interest, such as monotonicity. Although these theoretical properties are usually well understood in the sense that they are based on microeconomic theory, empirical researchers quite often overlook them. For example, production functions are monotonic in input quantities, yet when a translog production function is estimated, very rarely is monotonicity enforced *ex ante* or checked *ex post*. Imposition and testing of constraints is a key component of all empirical work; these constraints form the basis for the underlying economic theory producing the relationship of interest.

However, imposing constraints on empirical models presents its own challenge since it is traditionally more difficult to impose constraints on models which are more flexible whereas models easier to impose constraints may not be flexible enough to capture the underlying relationship of interest. In general, a tradeoff exists between estimating models which satisfy smoothness properties dictated by economic theory and models general enough to capture features of the unknown (and typically unspecified) relationship. The standard set of flexible functional forms (translog, Diewert, generalized McFadden, etc.) often violate theoretical properties given their capability of approximating a wide range of underlying relationships. Imposition of constraints is, in general, feasible, but might require the use of nonstandard methods or computational guile. Moreover, the results from alternative econometric methodologies (parametric vs. nonparametric, say) which ensure theoretical consistency differ markedly. There does not exist a simple, yet general methodology to impose smoothness constraints on econometric models across different estimation methodologies. Our aim in this paper is to fill this gap by extending a technique from the nonparametric regression literature known as constraint weighted bootstrapping (CWB) to the class of linear regression estimators.[1]

CWB has been used to impose monotonicity on an unspecified, univariate conditional mean (Hall & Huang 2001) and to impose various 'linear' constraints in a multivariate nonparametric setting (Du, Parmeter & Racine 2012).[2] However, the insights of this technique are not specific to estimation within nonparametric circles. We show that they hold for a much larger class of estimators, most notably, linear regression estimators. This suggests that the CWB method can be used in conjunction with standard parametric models to impose smoothness constraints. Typically, we are concerned with the parameter estimates stemming from parametric estimation. However, positivity, monotonicity and constant returns to scale, etc., are general restrictions on the overall smoothness of the functional relationship being estimated. The use of CWB to impose these conditions results in parameter estimates that are consistent with the desired economic theoretical properties.

---

[1]See Ajgaonkar (1965) for a formal definition of linear estimators.
[2]See Henderson & Parmeter (2009) for a survey of various methods (including this one) to impose constraints on nonparametrically estimated regression surfaces.

The virtue of using CWB is that it requires nothing more than simple quadratic programming techniques which are widely available in econometric software packages. Coupled with the ability to use the same technique to impose constraints across estimation methodologies signifies that this is a reliable and easily implementable method to impose smoothness constraints in empirical research.

Previous approaches to impose smoothness constraints parametrically include Diewert & Wales (1987) who propose methods for introducing curvature conditions globally in a cost function setting; Terrell (1996) who proposes Bayesian techniques to impose constraints at the observed data points (instead of globally); Rambaldi & Doran (1997) who suggest the singular value decomposition (SVD) to impose linear, time-varying constraints in a regression setting; and O'Donnell, Rambaldi & Doran (2001) who implement the SVD insights in a Bayesian framework. Related closely to our empirical context, O'Donnell & Coelli (2005) specifically discuss Bayesian estimation (incorporating smoothness constraints) of output distance functions.

In this paper we show how CWB can be applied to impose smoothness conditions on linear regression estimators. We then discuss a simple test for the conditions imposed. With these insights described, we proceed to estimate an input distance function both parametrically, using the translog functional form, and nonparametrically, using local-linear generalized kernel regression. Both models are estimated with and without the constraints imposed to form a basis of comparison. As we will discuss later, our empirical application compares both restricted and unrestricted estimates across econometric methodologies in a simple fashion. This opens the door for further empirical work to estimate and test models of economic theory.[3]

Our empirical setting involves estimation of an input distance function which is dual to a cost function. Thus, the input distance function satisfies a variety of smoothness conditions similar to that of the cost function. Since the functional form for the underlying technology is generally unknown, it is typical to employ a translog functional form to estimate an input distance function. The other alternative is to estimate it nonparametrically without assuming a functional form. The smoothness conditions on the estimated functions, whether parametric or nonparametric, need to be satisfied to be consistent with microeconomic theory. In empirical applications researchers

---

[3]The recent work of Sauer (2006) also advocated for checking and testing for theoretical consistency, albeit in a different setting.

often check whether these smoothness regularity conditions are satisfied at the mean of the data,[4] as opposed to being satisfied at every data point, and report results (scale economies, technical change, etc.) evaluated at the mean. This practice defeats the purpose of using micro data. Empirical findings may not be of much use for policy purposes if defined at the mean (which may not represent any producer in the sample), and theoretical restrictions are violated for many individual producers.

To showcase the method we consider a sufficiently general empirical application. Specifically, we use farm-level panel data from the Norwegian Farm Accountancy Survey to estimate a model (input distance function) with 4 outputs and 6 inputs both parametrically and nonparametrically. The results from our exercise show that both approaches produced varying degrees of violations of economic theoretical properties. For the translog model there were minimal violations in general, but one gradient had nearly 11% of the estimates of the wrong sign. The nonparametric estimator brought slightly increased percentages of violations. We used the constrained estimators to correct these violations. In addition, we compare the estimator on a host of dimensions. For example, we perform out-of-sample forecasts and show that the nonparametric model has better out-of-sample predictions, especially when we look at the constrained estimators.

The remainder of the paper is organized as follows. Section 2 describes the CWB approach for linear regression estimators. Section 3 discusses production theory and proposes an econometric model with constraints. Section 4 describes the data. Section 5 presents both unconstrained and constrained estimation results for both parametric and nonparametric models. Finally Section 6 concludes and offers several avenues for future research.

2. Constraint Weighted Bootstrapping for Linear Regression Estimators

We describe in detail the application of CWB for linear regression estimators. Previous work (Hall & Huang 2001) has focused on imposing constraints for nonparametric estimators using CWB. However, the method is not unique to nonparametric estimators and can be implemented in the more general framework of a linear estimator for a regression model. The idea involves re-weighting the response in a way that respects the theoretical constraints imposed by economic theory.

---

[4]This is not always a bad strategy as Ryan & Wales (2000) have shown that imposing a constraint at a single point may result in that constraint being satisfied at many points.

2.1. **Estimation.** To be precise, in what follows next, we need to introduce some notation. Let $\{Y_i, X_i\}_{i=1}^{n}$ denote sample pairs of response and explanatory variables where Y_i is a scalar, X_i is of dimension d, and n denotes the sample size. The goal is to estimate the unknown average response $g(\mathbf{x}) \equiv E(Y|X = \mathbf{x})$ subject to constraints on $g_j(\mathbf{x})$, where $j = 0, 1, \ldots, d$. For $j = 0$, $g_0(\mathbf{x})$ represents the function itself; for $j = 1, \ldots, d$, $g_j(\mathbf{x})$ represents the first order derivative of $g(\mathbf{x})$ with respect to the j-th continuous variable. In a parametric world, where $g(\mathbf{x})$ is specified *a priori,* we use the notation $g(\mathbf{x}; \beta)$.

We consider imposing arbitrary constraints on the estimate $\hat{g}(\mathbf{x})$ or $\hat{g}(\mathbf{x}; \beta)$ of the form

$$(1) \qquad\qquad l_j(\mathbf{x}) \leq \hat{g}_j(\mathbf{x}) \leq u_j(\mathbf{x})$$

for arbitrary $l_j(\cdot)$ and $u_j(\cdot)$, which represents (local) lower and upper bounds, respectively. For some applications, $l_j = 0$, $u_j = 1$, $\forall j \neq 0$, would be of particular interest, say, when the partial derivative represents a budget share and therefore must lie in $[0, 1]$. Or, $j = 0$ and $l_j = 0$, $u_j = 1$ might be of interest when an outcome must be bounded. For example, $\hat{g}_0(\mathbf{x}; \beta)$ could represent an estimate of the linear probability model and hence must lie in $[0, 1]$. Or, $l_j(\cdot) = u_j(\cdot)$ might be required (i.e., equality rather than inequality constraints) such as when imposing adding up constraints, say, when the sum of the budget shares must equal one, or when imposing homogeneity of a particular degree, by way of example. The approach we describe holds for general linear regression estimators. Until now, it has only been studied in strictly nonparametric frameworks (see Henderson & Parmeter 2009). It is firmly embedded in a conventional regression framework (parametric or nonparametric), and admits arbitrary combinations of constraints subject to the obvious caveat that the constraints must be internally consistent.

Consider a linear estimator of the regression model

$$(2) \qquad\qquad \hat{g}_j(\mathbf{x}) = \sum_{i=1}^{n} A_{j,i}(\mathbf{x}) Y_i,$$

where, without loss of generality, we have suppressed dependence on a parameter here for ease of discussion. In the subsections to come, we discuss construction of $A_{j,i}(\mathbf{x})$ for a variety of popular estimators. Specifically, we show how to construct the $A_{j,i}(\mathbf{x})$ components for the ordinary least-squares (OLS) estimator, as well as kernel-based nonparametric estimators.

The idea behind CWB is to recognize that the conditional mean estimator represents an equally weighted average. This is seen by rewriting the estimator in (2) as

$$(3) \qquad \hat{g}_j(\mathbf{x}) = n \sum_{i=1}^{n} A_{j,i}(\mathbf{x}) n^{-1} Y_i,$$

and letting $n^{-1} Y_i$ denote our response observation. We can see that each of the n terms in our average contribute equally. If, however, we accept that some terms are more important than others with respect to preserving theoretically consistent smoothness assumptions, then we can allow different points to contribute differently to the overall average. That is, by letting our weights differ from the uniform $1/n$ (but still requiring them to sum to one) that is inherently imposed in a general linear regression estimator, we can now manipulate the averaging in such a way that smoothness constraints can be imposed in a straightforward fashion. Thus, we still keep $n^{-1} Y_i$ as the unit of observation for the response, but we let our weighting depend on the index. This is done via the re-weighted estimator

$$(4) \qquad \hat{g}_j(\mathbf{x}|p) = n \sum_{i=1}^{n} A_{j,i}(\mathbf{x}) p_i Y_i,$$

where our uniform $1/n$ weighting mechanism has been replaced by observation specific weighting in the form of p_i. The weights are introduced to ensure that the constraints imposed on $\hat{g}_j(\mathbf{x})$ are preserved. A nice feature of this setup is that the smoothness of both $\hat{g}_j(\mathbf{x})$ and $\hat{g}_j(\mathbf{x}|p)$ depends on $\mathbf{x}$ through $A_{j,i}(\mathbf{x})$ so the estimator can also be considered, as linear in p.

It may appear that there is a conflict between the 'smoothness' of the estimator and the constraint weights. However, from our point of view these two issues are exclusive since we think of fixing the smoothness prior to imposing any constraints. This is true whether we are in a parametric or nonparametric setting given that one can think of selecting the functional form *a priori* as setting the level of smoothing in the parametric setting. Hall & Huang (2001) and Du et al. (2012) both require the bandwidth(s) for the nonparametric problem to have the optimal rate, irrespective of the constraint weights.

Linearity in p is crucial for ease of implementation. With linear in p constraints and a quadratic distance metric (to be discussed below), the constrained estimator can be implemented with

nothing more than quadratic programming. As noted by O'Donnell & Coelli (2005, pgs. 495-496) "...monotonicity constraints ... are easy to impose using linear programming, but difficult to impose using traditional econometric approaches, especially since they need to be imposed at each data point." Linear in p constraints include positivity, monotonicity, concavity,[5] supermodularity, symmetry, and adding up constraints on any order of a set of derivatives. When we use a distance metric that is not quadratic, or impose constraints that are not linear in p, we must resort to nonlinear programming methods, such as sequential quadratic programming (Henderson & Parmeter 2009, Section 3.1).

We now show how to impose particular restrictions on the estimator $\hat{g}_j(\mathbf{x}|p)$. Let p_u be an n-vector with elements $1/n$ and let p be the vector of weights to be selected. In order to impose our constraints, we choose p to minimize some distance measure from p to the uniform weights p_u as proposed by Hall & Huang (2001). This is appealing intuitively since the unconstrained estimator is that for which $p_i = 1/n \ \forall i$, as noted above. Whereas Hall & Huang (2001) consider probability weights (i.e., $0 \le p_i \le 1$, $\sum_i p_i = 1$) and distance measures suitable for probability weights (power divergence; Cressie & Read (1984)), it is possible to relax the constraint that $0 \le p_i \le 1$ and instead allow for both positive and negative weights (while retaining $\sum_i p_i = 1$).[6] This requires the use of an alternative distance metric, the well-worn L_2 metric

$$(5) \qquad\qquad D(p) = (p_u - p)'(p_u - p).$$

Our problem therefore boils down to selecting those weights p that minimize $D(p)$ subject to $l_j(\mathbf{x}) \le \hat{g}_j(\mathbf{x}|p) \le u_j(\mathbf{x})$ (and perhaps additional constraints of a similar form), which can be cast as a general nonlinear programming problem.[7] Constraints that are linear in p,[8] can be solved

---

[5]It is feasible (but difficult computationally) to impose concavity in a linear in p setup using the 'Afriat' conditions (Matzkin 1994); see Parmeter & Racine (2012) for details.

[6]To appreciate why it may prove necessary to allow for negative weighting, suppose we simply wished to constrain a surface that has both positive and negative regions to be uniformly positive. This could be accomplished by allowing some of the weights to be negative. However, probability weights would fail to produce a feasible solution as they are all non-negative.

[7]A referee has correctly noted the tight link between CWB and empirical likelihood. However, in that setup one minimizes $\sum_{i=1}^{n} \log(p_i/p_u)$ subject to a set of moment conditions. CWB as proposed here with a quadratic norm makes solving the constrained problem easier to implement and offers (potentially) extra flexibility to satisfy the constraints by allowing $p_i < 0$.

[8]Other useful economic constraints that satisfy (in)equalities that are linear in p include additive separability, homogeneity, diminishing marginal returns/products, and bounding of derivatives of any order.

using standard quadratic programming methods and off-the-shelf software. For example, in the R language (R Development Core Team 2008) it is solved using the *quadprog* package.[9] Even when n is quite large, the solution is computationally fast.

2.2. **Examples of Linear Regression Estimators.** Here we present several examples of linear regression estimators that fit into the framework of CWB. Consider a classical linear model and follow standard matrix notation: $Y = X\beta + u$ where Y is a response vector, X is a data matrix whose rows are X_i, β is a parameter vector and u is a disturbance. To estimate the model via OLS, we can write $A_{j,i}(\mathbf{x})$ in (2) as $A_{j,i}(\mathbf{x}) = \mathbf{x}_j(X'X)^{-1}X_i$, where $\mathbf{x}_j$, $j = 0, 1, \ldots, d$, is the value at which the function is to be evaluated. For $j = 0$, $\mathbf{x}_0$ consists of all the variables in a regression function; for $j = 1, \ldots, d$, $\mathbf{x}_j$ consists of the first order derivatives of each of the element in $\mathbf{x}_0$ with respect to the j-th continuous variable in $\mathbf{x}_0$.

Although they are often used to estimate nonlinear models, nonparametric kernels methods are linear regression estimators. Consider a nonparametric model: $Y_i = g(X_i) + u_i$, where $g(\cdot)$ is some unknown smooth function. To estimate the model using the local-constant kernel estimator, we can write $A_{0,i}(\mathbf{x})$ as

$$(6) \qquad A_{0,i}(\mathbf{x}) = \frac{K_\gamma(X_i, \mathbf{x})}{\sum_{i=1}^n K_\gamma(X_i, \mathbf{x})},$$

where $K_\gamma(\cdot)$ is a generalized product kernel that admits both continuous and categorical data, and γ is a vector of bandwidths (see Racine & Li (2004) for details).

To estimate the nonparametric model via the local-linear kernel estimator, which can most easily be thought of as fitting a line at each point of interest as opposed to fitting a constant (hence local-constant versus local-linear), we can write $A_{0,i}(\mathbf{x})$ as

$$
\begin{aligned}
A_{0,i}(\mathbf{x}) \;=\; & \left(\sum_{i=1}^n K_\gamma(X_i, \mathbf{x}) \begin{pmatrix} 1 & (X_i^c - \mathbf{x}^c) \\ (X_i^c - \mathbf{x}^c) & (X_i^c - \mathbf{x}^c)(X_i^c - \mathbf{x}^c)' \end{pmatrix} \right)^{-1} \\
& K_\gamma(X_i, \mathbf{x}) \begin{pmatrix} 1 \\ (X_i^c - \mathbf{x}^c) \end{pmatrix},
\end{aligned}
\tag{7}
$$

---

[9]Code in the R language is available from the authors upon request.

where X^c is a vector of continuous regressors. The local-linear estimator yields consistent estimates of the conditional mean as well as the d first-order derivatives of $g(\cdot)$ (Li & Racine 2007, Theorem 2.7).[10]

2.3. **Imposing a Simple Constraint.** To help the practitioner understand the estimator for a specific setting, we show how to construct the $A(\cdot)$ matrix for a particular example using both the OLS and nonparametric estimators. Suppose that we set $g(\mathbf{x}, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. To impose monotonicity in the parametric setting using OLS (at a particular point), we would need to ensure that $\hat{g}_1(\mathbf{x}, \beta|p) \geq 0$ and $\hat{g}_2(\mathbf{x}, \beta|p) \geq 0$. This is equivalent to imposing that $\beta_1 \geq 0$ and $\beta_2 \geq 0$. Our first order differentiated version of $A_{j,i}(\mathbf{x})$ would have

$$(8) \qquad A_{1,i}(\mathbf{x}) = \mathbf{x}_1(X'X)^{-1}X_i = (0,1,0)(X'X)^{-1}X_i$$

and

$$(9) \qquad A_{2,i}(\mathbf{x}) = \mathbf{x}_2(X'X)^{-1}X_i = (0,0,1)(X'X)^{-1}X_i.$$

If our unrestricted least-squares estimator produced slope coefficients that were already positive, then there would be no need to use the restricted estimator. However, if one or both of the estimates were negative, then the constraint weights would differ from the uniform weights to ensure the constraints were satisfied.

Alternatively, suppose that we set $g(\mathbf{x}, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. To impose monotonicity at a point, we would need to ensure that $\hat{g}_1(\mathbf{x}, \beta|p) \geq 0$. This is equivalent to imposing that $\beta_1 + 2\beta_2 x_1 \geq 0$. In constructing our first order differentiated version of A we would simply have

$$(10) \qquad A_{1,i}(\mathbf{x}) = \mathbf{x}_1(X'X)^{-1}X_i = (0,1,2x_1)(X'X)^{-1}X_i.$$

If we wished to impose monotonicity in the simple univariate setting nonparametrically then our first order differentiated version of $A_{1,i}(\mathbf{x})$ (assuming the use of a Gaussian kernel) would be,

---

[10]This methodology also works well with the general order local polynomial estimator as well as with spline regression. We leave a full treatment of this for future research.

recalling that in the local-constant least-squares setting $A_{0,i}(\mathbf{x}) = nK_i(\mathbf{x})/\sum_{i=1}^{n} K_i(\mathbf{x})$,

$$\frac{\partial A_{0,i}(\mathbf{x})}{\partial \mathbf{x}} = A_{1,i}(\mathbf{x}) = \frac{n\frac{\partial K_i(\mathbf{x})}{\partial \mathbf{x}}\sum_{i=1}^{n} K_i(\mathbf{x}) - nK_i(\mathbf{x})\sum_{i=1}^{n}\frac{\partial K_i(\mathbf{x})}{\partial \mathbf{x}}}{\left[\sum_{i=1}^{n} K_i(\mathbf{x})\right]^2}$$

$$(11) \qquad\qquad = A_{0,i}(\mathbf{x})\left[n^{-1}\sum_{i=1}^{n} D_i(\mathbf{x}_s)A_{0,i}(\mathbf{x}) - D_i(\mathbf{x}_s)\right] = A_{0,i}(\mathbf{x})M_s(\mathbf{x}),$$

where $D_i(\mathbf{x}_s) = \left(\frac{X_{1i}^c - x_1^c}{\gamma_{c1}^2}, \frac{X_{2i}^c - x_2^c}{\gamma_{c2}^2}, \cdots, \frac{X_{di}^c - x_d^c}{\gamma_{cd}^2}\right)'$ is a vector where the c sub/superscript signifies only the continuous variables (since we are differentiating).

2.4. **Graphical Illustration.** To further cement these ideas we provide a simple graphical illustration on how CWB works. Suppose that we wished to impose monotonicity on a conditional mean, but our data are generated from $Y = X_1^3 + \varepsilon$ which is monotonic on the domain of interest, $[-2,2]$. However, for a draw of 100 observations with $X_1 \sim \mathcal{U}[-2,2]$ and $\varepsilon \sim N(0,1)$, both traditional OLS and the local-constant estimator produce conditional means which are non-monotonic. The bandwidth for the local-constant estimator is selected via AIC_c (Hurvich, Simonoff & Tsai 1998) and is used for both the unrestricted and restricted smooths. The parametric estimates are generated by fitting a third order polynomial in the covariate. The restricted and unrestricted function estimates are displayed in Figure 1. The parametric estimates appear in panel (a) while the nonparametric estimates appear in panel (b).

We see that the noise induced from this particular sample has obscured the monotonic relationship. Both methods successfully mimic the monotonic behavior of the data generating process when they are restricted via constraint weighted bootstrapping. To appreciate how the averaging is re-weighted in order to constrain the estimator, Figure 2 plots out the original points against the constraint weighted points for each estimator. This shifting of the unconstrained datum now allows the fit of the function to preserve monotonicity. The idea is that the data is moved as little as possible to uphold the constraint(s) imposed. The determination of how little to move the weights from their uniform counterparts is dictated by the choice of distance metric.[11] Notice that the

---

[11]Both Hall & Huang (2001) and Henderson & Parmeter (2009) argue that for the most part, the choice of distance metric has relatively little influence on the weights and the resulting estimates.

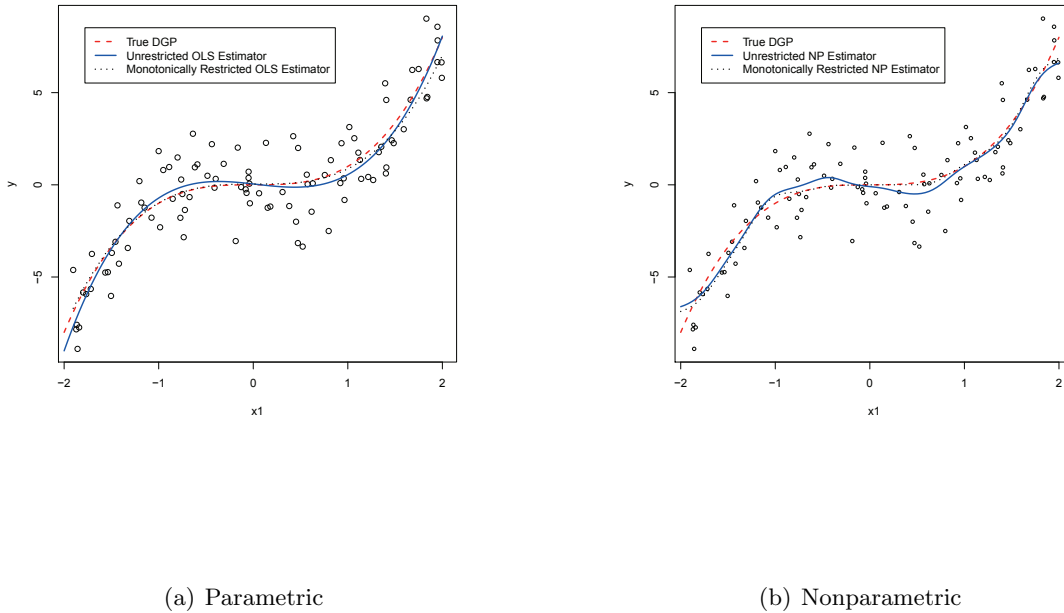(a) Parametric   (b) Nonparametric

FIGURE 1. Restricted and unrestricted conditional mean estimates of $Y = X_1^3 + \varepsilon$ where monotonicity is imposed, i.e., $j = 1$, $\hat{g}_1(\mathbf{x}|p) \geq 0$, $n = 100$.

reweighted values of Y_i differ across the two estimators. For the parametric estimator it appears that points towards the end of the support of the data are reweighted while for the local-constant estimator, it appears that points closer towards the middle of the support are reweighted. This underscores the difference between a global parametric estimator and a local nonparametric estimator. Both restricted estimators impose monotonicity on the conditional mean, but they reweight the data in entirely different manners.

2.5. **Inference.** There exists a growing literature on testing restrictions. For example, Abrevaya & Jiang (2005) test for curvature restrictions, Epstein & Yatchew (1985) develop a nonparametric test of the utility maximization hypothesis and homotheticity, and Yatchew & Bos (1997) develop a conditional moment test for a broad range of smoothness constraints. Aside from the test of Yatchew & Bos (1997), most existing tests check for specific constraints. This is limiting in the current setting as our main focus is on a smooth, arbitrarily restricted estimator.

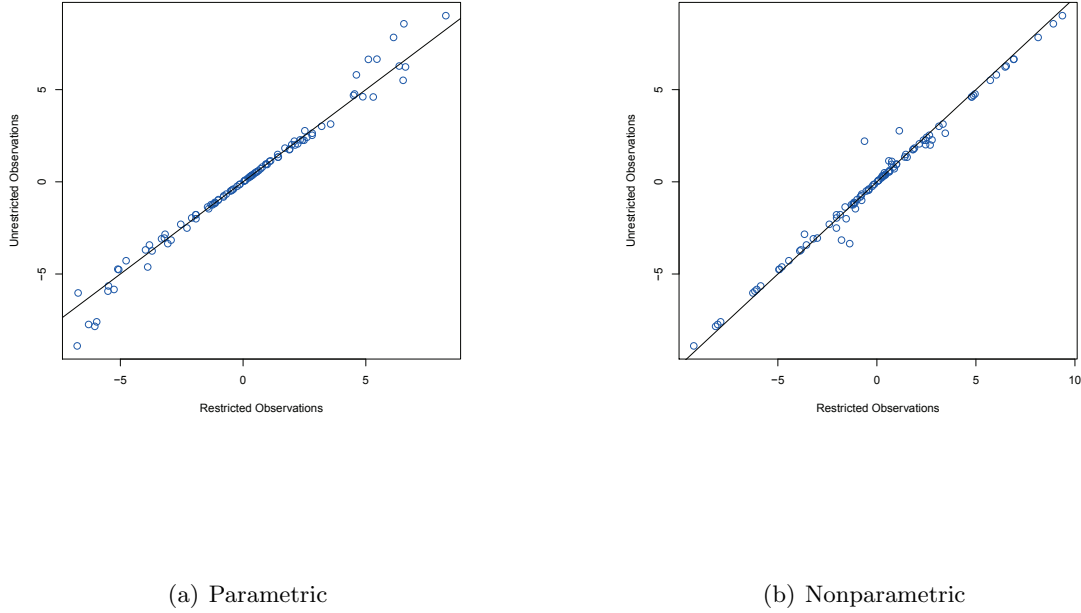(a) Parametric   (b) Nonparametric

FIGURE 2. Restricted and unrestricted points used for estimation in Figure 1.

Here we adopt a testing approach similar to that initially proposed by Hall, Huang, Gifford & Gijbels (2001) which is predicated on the objective function $D(\hat{p})$. This approach involves estimating the constrained regression function $\hat{g}(\mathbf{x}|p)$, based on the sample realizations $\{Y_i, X_i\}$, and then rejecting H_0 that constraints are valid, if the observed value of $D(\hat{p})$ is too large. We use a re-sampling approach for generating the null distribution of $D(\hat{p})$ which involves generating re-samples for y drawn from the constrained model via wild residual re-sampling (Härdle & Mammen 1993), which is robust to heteroskedasticity. These re-samples are generated under H_0, hence we recompute $\hat{g}(\mathbf{x}|p)$ for the bootstrap sample $\{Y_i^*, X_i\}$, and obtain $\hat{g}(\mathbf{x}|p^*)$, which then yields $D(p^*)$. We then repeat this process a large number (B) of times. Finally, we compute the empirical p-value, P_B, which is simply the proportion of the B bootstrap re-samples $D(p^*)$ that exceed $D(\hat{p})$, i.e.,

$$(12) \qquad P_B = 1 - \hat{F}(D(\hat{p})) = \frac{1}{B}\sum_{b=1}^{B} 1(D(p^*) > D(\hat{p})),$$

where $1(\cdot)$ is the indicator function and $\hat{F}(D(\hat{p}))$ is the empirical distribution function of the bootstrap statistics. We can reject the null hypothesis if P_B is less than α, the level of the test.

An outline of the testing procedure is as follows:

(1) Estimate the restricted model and obtain $D(\hat{p})$.

(2) For $i = 1, 2, \ldots, n$, generate the two-point wild bootstrap error $\hat{\varepsilon}_i^* = \left[\left(1 - \sqrt{5}\right)/2\right]\hat{\varepsilon}_i$, where $\hat{\varepsilon}_i = y_i - \hat{g}(\mathbf{x}_i|p)$ with probability $r = \left(1 - \sqrt{5}\right)/2\sqrt{5}$ and $\hat{\varepsilon}_i^* = \left[\left(1 + \sqrt{5}\right)/2\right]\hat{\varepsilon}_i$ with probability $1 - r$. Construct a wild estimate for Y using the original restricted regression estimates (fitted values) and the re-sampled wild residuals. Call these new values Y^*.

(3) Reestimate the restricted regression using the pairs $\{Y_i^*, X_i\}$ and obtain a new set of constraint weights, $\hat{p}^*$. Calculate and store the value $D(\hat{p}^*)$.

(4) Repeat steps 2 and 3 a large (B) number of times.

(5) Calculate the bootstrapped p-value of the test statistic as in (12).

Before proceeding further, note that there exist three situations that can occur in practice: (1) impose non-binding constraints (they are 'correct' de facto); (2) impose binding constraints that are correct; (3) impose binding constraints that are incorrect. For the implementation of the test just described, we only need to consider situations 2 and 3 since, as noted by Hall et al. (2001, p. 609), "For those datasets with $D(\hat{p}) = 0$, no further bootstrapping is necessary $[\ldots]$ and so the conclusion (for that dataset) must be to not reject H_0." The implication is simply that imposing non-binding constraints does not alter the estimator and the constrained weights will still be $\hat{p}_i = 1/n \;\forall i$, hence $D(\hat{p}) = 0$ and the statistic is degenerate. Of course, in practice this simply means that we presume people are imposing constraints that bind, which is a reasonable presumption. However, due to inappropriate bandwidth selection, the presence of outliers, measurement errors in the data, etc., the estimated function may not immediately satisfy the theoretically consistent constraints. Thus, in these cases testing will be warranted.

3. PRODUCTION THEORY AND EMPIRICAL MODELS

In this section we provide an empirical application of our proposed methodology. Specifically, we use Norwegian farm level dairy data to estimate an input distance function with 6 inputs and 4 outputs. Thus, our starting point is to specify a multi-output multi-input flexible transformation

function and use appropriate normalizing constraints (normalizations) which gives the input distance function. Our departure from the usual specification of a distance function is that we do not have to introduce inefficiency into our model; see Shephard (1953, 1970) and McFadden (1978), among others. Input and output distance functions offer an attractive alternative towards the estimation of multi-input, multi-output production technologies without the necessity of explicitly introducing behavioral assumptions of producers, and without using information on prices which are often difficult to obtain. Consistent with economic theory, we constrain the gradients of nine regressors simultaneously. Specifically, for each observation, we require that the gradients on the inputs be non-positive and the gradients on the outputs to be non-negative.

In what follows, we first derive the input distance function starting from the transformation function in microeconomic theory. This provides the basis for our empirical econometric model (the input distance function). We then discuss the requisite economic theoretical constraints based upon our use of the input distance function.

3.1. **Transformation and Input Distance Functions.** We assume that a producer uses Q inputs, X, to produce a vector of M outputs, Y.[12] The functional relationship between X and Y is expressed as $f(Y, X, t) = 1$, where $f(Y, X, t)$ is the transformation function, and t represents technology proxied by the time trend variable. Since distance functions are mostly used for estimating inefficiency and our focus in this paper is not on efficiency, we start from the transformation function representation of the production technology and show that the input and output distance functions (plus some other formulations) can be derived from it by using appropriate normalizations (Kumbhakar 2012).

If we assume that the transformation function is translog (TL), then it can be expressed as

$$\ln f(Y, X, t) = \alpha_0 + \alpha_t t + \frac{1}{2}\alpha_{tt}t^2 + \sum_{m=1}^{M} \alpha_m \ln Y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{l=1}^{M} \alpha_{ml} \ln Y_m \ln Y_l$$

$$+ \sum_{q=1}^{Q} \beta_q \ln X_q + \frac{1}{2}\sum_{q=1}^{Q}\sum_{o=1}^{Q} \beta_{qo} \ln X_q \ln X_o + \sum_{m=1}^{M}\sum_{q=1}^{Q} \delta_{mq} \ln Y_m \ln X_q$$

$$\text{(13)} \qquad + \sum_{m=1}^{M} \alpha_{mt} \ln Y_m t + \sum_{q=1}^{Q} \beta_{qt} \ln X_q t,$$

---

[12]Note that X and Y here does not denote regressor and regressand, respectively.

where $\beta_{qo} = \beta_{oq}$ and $\alpha_{ml} = \alpha_{lm} \ \forall q, o, m$ and l. If we rewrite (13) as

$$\ln f(Y, X, t) = \alpha_0 + \alpha_t t + \frac{1}{2}\alpha_{tt}t^2 + \sum_{m=1}^{M} \alpha_m \ln Y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{l=1}^{M} \alpha_{ml} \ln Y_m \ln Y_l + \sum_{q=2}^{Q} \beta_q \ln(X_q/X_1)$$

$$+ \frac{1}{2}\sum_{q=2}^{Q}\sum_{o=2}^{Q} \beta_{qo} \ln(X_q/X_1)\ln(X_o/X_1) + \sum_{m=1}^{M}\sum_{q=2}^{Q} \delta_{mq} \ln Y_m \ln(X_q/X_1)$$

$$+ \left[\sum_{q=1}^{Q}\beta_q\right] \ln X_1 + \sum_{q=1}^{Q}\left[\sum_{o=1}^{Q}\beta_{qo}\right] \ln X_q \ln X_1 + \sum_{m=1}^{M}\left[\sum_{q=1}^{Q}\delta_{mq}\right] \ln Y_m \ln X_1$$

$$(14) \qquad + \sum_{m=1}^{M} \alpha_{mt} \ln Y_m t + \sum_{q=2}^{Q} \beta_{qt} \ln(X_q/X_1)t + \left[\sum_{q=1}^{Q}\beta_{qt}\right] \ln X_1 t,$$

and use the following normalizations $\sum_q \beta_q = -1$, $\sum_o \beta_{qo} = 0 \ \forall q$, $\sum_q \delta_{mq} = 0 \ \forall m$, and $\sum_q \beta_{qt} = 0$, we get the input distance function representation, viz.,

$$\ln X_1 = \alpha_0 + \alpha_t t + \frac{1}{2}\alpha_{tt}t^2 + \sum_{q=2}^{Q} \beta_q \ln \tilde{X}_q + \frac{1}{2}\sum_{q=2}^{Q}\sum_{o=2}^{Q} \beta_{qo} \ln \tilde{X}_q \ln \tilde{X}_o$$

$$+ \sum_{m=1}^{M} \alpha_m \ln Y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{l=1}^{M} \alpha_{ml} \ln Y_m \ln Y_l + \sum_{m=1}^{M}\sum_{q=2}^{Q} \delta_{mq} \ln Y_m \ln \tilde{X}_q$$

$$(15) \qquad + \sum_{m=1}^{M} \alpha_{mt} \ln Y_m t + \sum_{q=2}^{Q} \beta_{qt} \ln \tilde{X}_q t,$$

where $\tilde{X}_q = X_q/X_1, q = 2, \cdots, Q$. Note that the above normalizations make the transformation function in (13) homogenous of degree 1 in X_1 (Shephard 1953), and gives it an IDF representation in (15).

Consequently, instead of discussing the properties of the transformation function, we can simply refer to the properties of the input distance function (among others see Shephard 1953, McFadden 1978, Färe & Primont 1995, Kumbhakar & Lovell 2000). The properties of the IDF are: (i) $f(0, X, t) = +\infty$, $f(Y, 0, t) = 0$, (ii) $f(Y, X, t)$ is homogeneous of degree 1 in X, (iii) $f(Y, X, t)$ is non-decreasing in X (i.e., $f(Y, \lambda X, t) \geq f(Y, X, t)$ for $\lambda \geq 1$), (iv) $f(\lambda Y, X, t) \leq f(Y, X, t)$ for $\lambda \geq 1$, (v) $f(Y, X, t)$ is concave in X, and (vi) $f(Y, X, t)$ is continuous. The properties in (iii) and (iv) are monotonicity properties which along with the homogeneity property in (ii) will be our main focus in the paper. These properties are similar to the properties of the cost function when X is

the vector of input prices. Thus, the above properties of the IDF can also be derived from the properties of a cost function.[13]

3.2. **Econometric Model and Economic Constraints.** Instead of starting from a parametric transformation function, we can start from the implicit form, i.e., $f(Y, X, t) = 1$ and impose the homogeneity property (ii) to write it as

$$(16) \qquad \ln X_1 = r(\ln \tilde{X}_2, \ldots, \ln \tilde{X}_Q, \ln Y_1, \ldots, \ln Y_M, t),$$

where $\tilde{X}_q = X_q/X_1$, $q = 2, \ldots, Q$. Here the first input (X_1) is used as a numeraire to impose the linear homogeneity property in the transformation function. It does not matter which one of the inputs is chosen as the numeraire (i.e., the results are invariant to the choice of numeraire). We estimate $r(\cdot)$ in (16) using the kernel-based local-linear nonparametric method. Alternatively, since (16) holds, all we need to do is assume a functional form for $r(\cdot)$ for the parametric model. Again note that we do not need to assume the presence of inefficiency in the production process to obtain the specification in (15).

The monotonicity conditions in (iii), based on the IDF in (15) are:

$$(17) \qquad \frac{\partial \ln X_1}{\partial \ln \tilde{X}_q} = \beta_q + \beta_{qt}t + \sum_{o=2}^{Q} \beta_{qo} \ln \tilde{X}_o + \sum_{m=1}^{M} \delta_{mq} \ln Y_m \leq 0, \ q = 2, \ldots, Q.$$

Using the linear homogeneity (in X) conditions, we get

$$\frac{\partial \ln f(Y, X)}{\partial \ln X_1} = 1 - \sum_{q=2}^{Q} \frac{\partial \ln f(Y, X)}{\partial \ln X_q} = 1 - \sum_{q=2}^{Q} \frac{\partial \ln X_1}{\partial \ln \tilde{X}_q} = 1 - \sum_{q=2}^{Q} \left[\beta_q + \sum_{o=2}^{Q} \beta_{qo} \ln \tilde{X}_o + \sum_{m=1}^{M} \delta_{mq} \ln Y_m \right].$$

The monotonicity conditions in (iv), based on the IDF in (15), are :

$$(18) \qquad \frac{\partial \ln X_1}{\partial \ln Y_m} = \alpha_m + \alpha_{mt}t + \sum_{l=1}^{M} \alpha_{ml} \ln Y_l + \sum_{q=2}^{Q} \delta_{mq} \ln \tilde{X}_q \geq 0, \ m = 1, \ldots, M.$$

When we employ local-linear generalized kernel estimation as discussed in Section 2.2, we are able to simultaneously obtain the conditional mean estimate, along with first order derivatives, $\partial \ln X_1/\partial \ln \tilde{X}_q$, $\forall q = 2, \ldots, Q$, and $\partial \ln X_1/\partial \ln Y_m$, $\forall m = 1, \ldots, M$, which are also observation-specific, and need to jointly satisfy the inequality constraints in (17) and (18). This results in

---

[13]See Färe & Primont (1995) and Kumbhakar & Lovell (2000) for more details on the properties of distance functions.

$Q + M - 1$ monotonicity constraints being imposed simultaneously on each observation for either the parametric or nonparametric input distance function. To control for individual effects, we can also estimate a fixed effect model by putting a categorical individual id variable into $r(\cdot)$. In the translog case, this variable becomes a set of dummy variables, while in the nonparametric case, it is simply a single unordered discrete variable which can be smoothed through a kernel function (see Racine & Li (2004) for details).

4. DATA

Our data come from the Norwegian Farm Accountancy Survey. This is a farm-level panel data collected by the Norsk institutt for landbruksøkonomisk forskning (NILF), an independent research institute under the Ministry of Agriculture in Norway. It includes farm production and economic data collected annually from about 1000 farms from different regions, farm size classes, and types of farms. Participation in the survey is voluntary. There is no limit on the number of years a farm may be included in the survey. Approximately 10% of the farms surveyed are replaced every year.

The farms are classified according to their main category of farming, defined in terms of the standard gross margins of the farm enterprises. For example, the main share of the total standard gross margin for farms categorized as dairy farms comes from dairy production. The dataset used in this analysis is a large unbalanced panel with 4333 observations on 714 dairy farms observed during 1993-2006. Dairy farms are usually involved in other farm production activities such as beef, grain, machinery contracting work, etc.

We distinguish between four outputs: milk sold measured in liters (Y_1), beef and cattle produced (Y_2), support payments (Y_3), and other outputs (Y_4). The last three outputs are converted from nominal monetary values to real (2006 Norwegian Kroners – NOK) values. The variable beef and cattle produced is also converted to 2006 values using the price index for cattle; the variable support payments is converted using the consumer price index (CPI); and the variable other outputs is deflated by the price index for other agricultural production. The beef and cattle production comprises both sales of cattle for slaughter and livestock and the net increase of the livestock of cattle. The support payments comprise various (more or less) decoupled farm supports. Other

outputs represent farm products other than milk, beef, and cattle. Examples of other outputs are livestock products such as pigs, sheep, goats, and crop products.

Our inputs are comprised of land measured in decares (1.0 decares = 0.1 hectare) (X_1), own and hired farm labor in hours (X_2), purchased feed (X_3), other variable costs (X_4), cattle capital (X_5), and other capital (X_6). The last four inputs are all measured in 2006 NOK. Purchased feed is deflated by the purchased feed price index, other variable costs is converted to 2006 values by the price index for other variable costs, cattle capital is deflated by price index for cattle, and other capital is deflated by the price index for other agricultural production. Real outputs and inputs prices are constructed by deflating aggregate accounts in nominal NOK with suitable price indices. The price information is taken from the farm survey when available, and when not, from the agricultural sector of the national accounts. The second source of price information is countrywide price indices. Prices from the survey are aggregate values divided by aggregate quantities in certain regions. The regions are NUTS4 and NUTS3 (counties).[14]

Aggregation of prices is conducted in terms of Fisher indices (Diewert 1992). All prices are deflated to 2006 levels using the CPI published by Statistics Norway. An additional remark applies for the user cost of capital. The price indices of capital represent investment prices which are needed for deflation of capital stocks into real price measures. The user price is a proportion of the investment price: interest and depreciation. Both proportions are extracted from the farm account survey. Table 1 presents the descriptive statistics of the sample.

5. Estimation Results

The estimation results are given in Tables 5-4 and Figures 3-4. Table 5 presents the parameter estimates for both the constrained and unconstrained translog model. Tables 2 and 4 summarize the estimates of the gradient vectors for both the constrained and unconstrained parametric and nonparametric models, along with the percentage of observations which violate the economic constraints for each unconstrained estimator for each constraint as well as the p-value of testing the validity of the individual constraint. Table 3 reports the bandwidths used in the estimation of the nonparametric model. Finally, in addition to the simple summary of gradients, Figures 3 and 4

---

[14]Nomenclature of Territorial Units for Statistics (NUTS) is a geocode standard, developed and regulated by the European Union, for referencing the subdivisions of countries for statistical purposes.

TABLE 1. **Summary Statistics of the Variables**

Symbol	Variable Name	Mean	Sd.	Min.	Max.
X_1	Land (decares)	209.8	89.2	40.0	685.0
X_2	Labor (hours)	4.254	1.274	0.797	11.657
X_3	Purchased Feed	129.7	60.8	3.0	533.7
X_4	Other Inputs	242.2	96.7	35.2	919.9
X_5	Cattle Capital	214.8	96.8	20.1	939.6
X_6	Other Capital	1120.7	614.2	69.3	6692.0
Y_1	Milk (Liters)	84.8	33.3	10.2	273.5
Y_2	Beef and Cattle	137.3	86.2	0.0	1028.4
Y_3	Support Payments	266.9	51.0	70.5	520.5
Y_4	Other Outputs	47.7	66.4	0.0	1162.7

The variables are output and input indices.

plot the kernel density functions of the gradient estimates for the unconstrained and constrained parametric and nonparametric models, respectively.

5.1. **Translog.**

5.1.1. *Unconstrained Estimator.* In Table 2 we analyze the gradients for the translog model.[15] Specifically, we report the minimum, first quartile, median, mean, third quartile, and maximum values, labeled as min, Q1, median, mean, Q3 and max, respectively, for each of the gradient vectors along with their corresponding (wild) bootstrapped standard errors. It is promising to see that each of the quartile values take their expected sign. However, not all observations take their expected signs.

In order to get an idea of how many observations have estimated gradients which are of the wrong sign, Table 2 also gives the percentage of violations for each of the gradients in the unrestricted model. We generally do not observe many violations for each individual gradient and in one case we have none. However, the final column shows that 10.52% of the gradient estimates for $\ln X_4$ are of the wrong sign. What is interesting is that there is nothing from Table 1 that would suggest this variable is ill-behaved. This provides the motivation to consider a constrained estimator.

---

[15]See Appendix A for the estimates of the 65 parameters from our translog model.

TABLE 2. **Summary Statistics of Partial Derivatives: Translog**

	$\frac{\partial \ln X_1}{\partial \ln X_2}$	$\frac{\partial \ln X_1}{\partial \ln X_3}$	$\frac{\partial \ln X_1}{\partial \ln X_4}$	$\frac{\partial \ln X_1}{\partial \ln X_5}$	$\frac{\partial \ln X_1}{\partial \ln X_6}$	$\frac{\partial \ln X_1}{\partial \ln Y_1}$	$\frac{\partial \ln X_1}{\partial \ln Y_2}$	$\frac{\partial \ln X_1}{\partial \ln Y_3}$	$\frac{\partial \ln X_1}{\partial \ln Y_4}$
Unrestricted									
Min	-0.4182	-0.3015	-0.2592	-0.5118	-0.1449	-0.1810	-0.1521	-0.0860	-0.0307
	(0.0548)	(0.0238)	(0.0276)	(0.0766)	(0.0154)	(0.0120)	(0.0415)	(0.0111)	(0.0069)
Q1	-0.1920	-0.1692	-0.1337	-0.2884	-0.0733	0.2519	0.0992	0.1343	0.0065
	(0.0094)	(0.0120)	(0.0086)	(0.0058)	(0.0050)	(0.0008)	(0.0003)	(0.0007)	(0.0001)
Median	-0.1637	-0.1453	-0.1122	-0.2695	-0.0581	0.3131	0.1202	0.1894	0.0133
	(0.0007)	(0.0005)	(0.0004)	(0.0003)	(0.0003)	(0.0013)	(0.0004)	(0.0011)	(0.0001)
Mean	-0.1641	-0.1435	-0.1117	-0.2702	-0.0578	0.3138	0.1188	0.1890	0.0128
	(0.0026)	(0.0024)	(0.0017)	(0.0041)	(0.0010)	(0.0050)	(0.0018)	(0.0032)	(0.0002)
Q3	-0.1338	-0.1191	-0.0907	-0.2510	-0.0426	0.3785	0.1405	0.2443	0.0195
	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0002)	(0.0289)	(0.0092)	(0.0200)	(0.0006)
Max	0.0446	0.1816	0.0544	-0.0986	0.0654	0.7473	0.2620	0.5390	0.0467
	(0.0140)	(0.0427)	(0.0067)	(0.0069)	(0.0077)	(0.0509)	(0.0222)	(0.0669)	(0.0044)
Pct. Of Violation	0.12%	0.44%	0.25%	0.00%	0.42%	0.30%	0.44%	1.11%	10.52%
Restricted									
Min	-0.3457	-0.2209	-0.2325	-0.3642	-0.1279	0.0000	0.0000	0.0000	0.0000
	(0.0232)	(0.0147)	(0.0179)	(0.0175)	(0.0109)	(0.0018)	(0.0010)	(0.0061)	(0.0003)
Q1	-0.1956	-0.1465	-0.1346	-0.2815	-0.0717	0.2616	0.0932	0.1496	0.0084
	(0.0090)	(0.0057)	(0.0066)	(0.0167)	(0.0036)	(0.0006)	(0.0002)	(0.0006)	(0.0000)
Median	-0.1720	-0.1333	-0.1166	-0.2667	-0.0577	0.3065	0.1105	0.1935	0.0107
	(0.0006)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0010)	(0.0004)	(0.0009)	(0.0000)
Mean	-0.1722	-0.1322	-0.1162	-0.2668	-0.0574	0.3059	0.1089	0.1925	0.0107
	(0.0028)	(0.0020)	(0.0018)	(0.0040)	(0.0009)	(0.0048)	(0.0017)	(0.0032)	(0.0002)
Q3	-0.1469	-0.1187	-0.0991	-0.2524	-0.0426	0.3523	0.1259	0.2348	0.0130
	(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0219)	(0.0093)	(0.0154)	(0.0008)
Max	0.0000	0.0000	0.0000	-0.1726	0.0000	0.5686	0.1933	0.4328	0.0225
	(0.0033)	(0.0044)	(0.0025)	(0.0027)	(0.0024)	(0.0227)	(0.0064)	(0.0338)	(0.0014)
p-value	0.3636	0.0000	0.4040	1.0000	0.0548	0.0101	0.0000	0.2828	0.0000

1. The numbers in the parentheses are standard errors. Standard errors are calculated using a wild bootstrap with 99 replications.
2. The *p*-value is the probability of rejecting the null that the constraint is valid when it is true.

To view the entire set of estimates for the unconstrained translog model, we plot the kernel density estimates of the gradients in Figure 3. The dashed density in each panel is the kernel density estimate for the translog model without constraints. A vertical line is drawn at zero for the plots without constraints. The shaded area highlights the violations in the unconstrained models. We can clearly see that the density for the gradients of $\ln Y_4$ crosses into the negative region.

Although the signs of the gradients are relatively well behaved, we should also look at other measures of interest. One measure of economic importance is returns to scale. Specifically, in this context, the returns to scale (RTS) is calculated as the reciprocal of $\sum_{m=1}^{4} \frac{\partial \ln X_1}{\partial \ln Y_m}$. For our unconstrained estimator, the median RTS is 1.5701 with standard error 0.0024. This value is relatively high as we would typically expect a RTS in the neighborhood of unity. Given that we are adding together negative gradient estimates the estimated RTS is low. The constrained model will likely change this estimate.

5.1.2. *Constrained Estimator.* Given that we have some violations of economic theory, we consider constrained estimation of our translog function. The parameter estimates of the constrained translog function are also given in Table 5.[16] Here, 25 slope coefficients are significant at the 10% level. The signs of the significant coefficients are the same across the unconstrained and constrained models. The goodness of fit of this model is also high, with the constrained estimator having a squared correlation coefficient between the actual and fitted values are the same (as the unconstrained model) to four decimal places.

Again, with this many coefficients, it is difficult to examine economic relevance. Therefore we present the gradients in Table 2 and Figure 3. It is obvious from Figure 3 that the constrained estimators have smaller spread than the unconstrained counterpart for each gradient.[17] Note that the estimates that were of the wrong sign are not simply set to zero. There also does not appear to be a uniform shift in the estimates. It is also important to note that even in the cases where none of the gradients violate economic theory ($\ln \tilde{X}_5$), there are still minor changes in the gradients. This is because in the translog model the gradients are functions of coefficients which need to be augmented in order to satisfy separate constraints. This typically impacts every gradient estimate, even those which are not violated.

Now that all of our gradients have the appropriate sign, we reconsider the problem associated with RTS. One problem in the unconstrained case is that we were summing gradients some of which

---

[16]Our unbalanced panel includes $n \equiv \sum_{i=1}^{N} T_i = 4333$ total observations and we want to impose nine inequality constraints. In cases such as this, a relatively powerful computer in terms of speed and memory space is needed. For example, with a single constraint, the linear regression estimator is expressed as a $n \times n$ dimensional matrix. Our problem required three gigabytes of memory for the parametric model and four gigabytes for the nonparametric model.

[17]We use the Silverman reflection method to construct the densities for the constrained gradients.

FIGURE 3. **Gradient estimates from the unconstrained and constrained translog model**
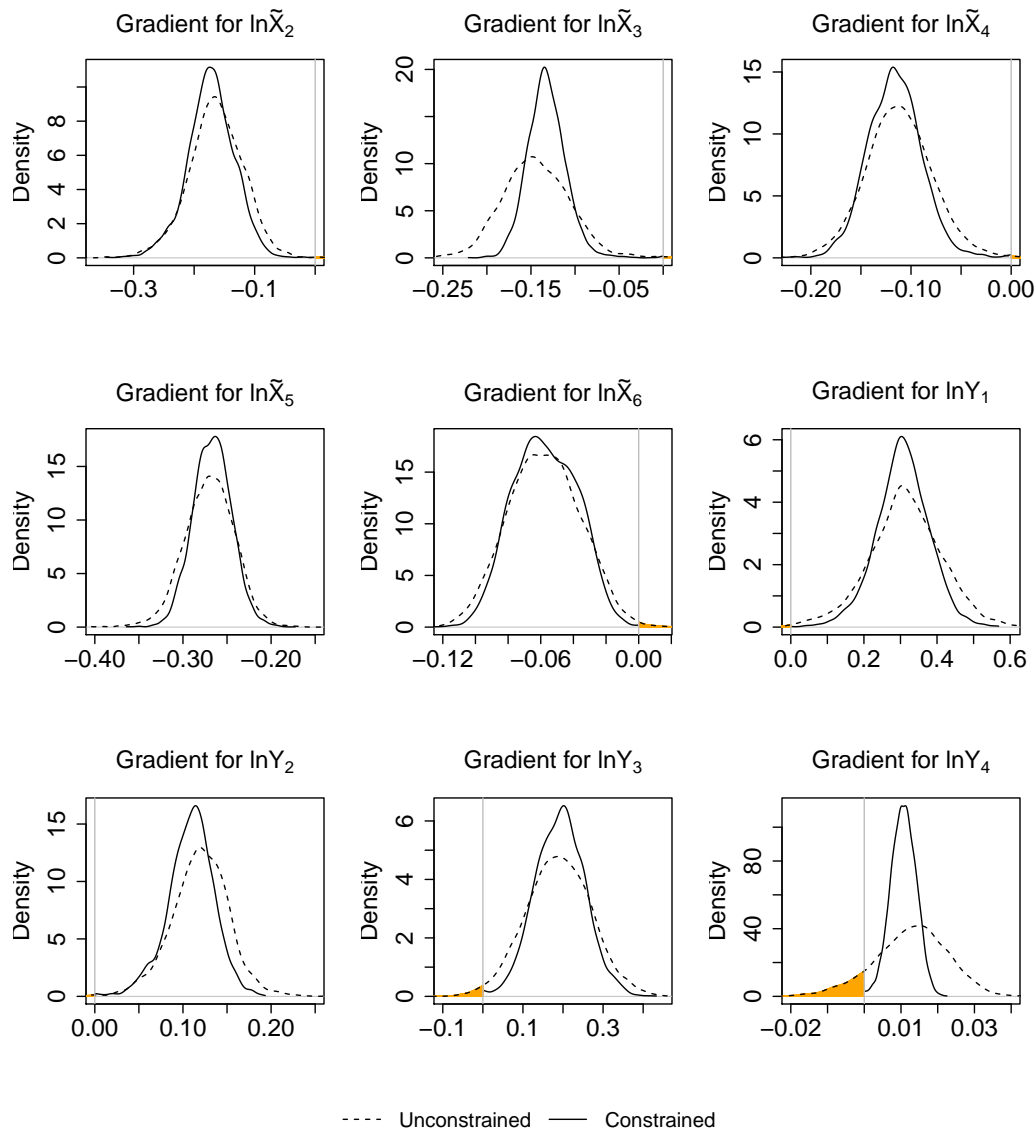
were negative. With the constrained estimator we do not have this particular problem. However, the median RTS in the constrained model is still relatively high (1.6131 with standard error 0.0022).

In summary, the translog function produced estimates which had violations in all but one gradient. The number of violations in one case was nearly 11%. The constrained estimator fixed these

violations. Considering this as well as possible functional form misspecification, it is worthwhile to consider nonparametric methods.

5.2. **Nonparametric.**

5.2.1. *Unconstrained Estimator.* Before rushing to the gradient estimates, we first examine the bandwidths from the local-linear least-squares procedure. Recent research has shown that cross-validation procedures can remove irrelevant variables as well as detect linearity. Here the band-widths are not simply nuisance parameters, but they can be helpful in uncovering the true underlying structure of the data. For example, the bandwidth on a categorical variable goes between 0 and 1. If the estimate is 1, then the variable is deemed irrelevant in the prediction of the left-hand-side variable as all observations are given equal weight. When the bandwidth is 0, it effectively means that the sample is being split based on that particular category. If the categorical variable is a typical dummy variable, and its bandwidth is equal to zero, in effect, two separate regressions are being estimated. In the case of a continuous regressor, the bandwidths go from zero to infinity. A bandwidth of infinity (or a sufficiently large value in practice) means that a variable enters in linearly. However, this linearity does not imply separability and hence heterogeneity in the gradients may still arise.

For both our constrained and unconstrained local-linear least-squares estimators, all bandwidths are selected via least-squares cross-validation criteria. The bandwidths are given in Table 3. Again, we have one categorical variable (farm id) and ten continuous variables. The most striking feature is that the bandwidth on the categorical variable is relatively small. This means that we are putting less weight on other farms to estimate the result for a particular farm. In terms of the continuous regressors, the bandwidths on $\ln \tilde{X}_2$, $\ln Y_m, m = 1, \ldots, 4$ and t are less than two times their respective standard deviations (rule of thumb for linearity) and hence they enter in nonlinearly. On the other hand, the bandwidths for $\ln \tilde{X}_5$ and $\ln \tilde{X}_6$ are slightly larger than two times their respective standard deviations while $\ln \tilde{X}_5$ and $\ln \tilde{X}_6$ have very large bandwidths. It can be argued (to different degrees) that these variables enter the model linearly. We want to note that nonlinearity does not exclude the translog model as it allows for quadratics and interactions. We will perform a formal test for correct parametric specification later.

TABLE 3. **Bandwidth Summary**

Variable	Bandwidth	Upper Bound	Interpretation
Continuous			
$\ln \tilde{X}_2$	0.575	$+\infty$	nonlinear
$\ln \tilde{X}_3$	0.916	$+\infty$	likely linear
$\ln \tilde{X}_4$	0.758	$+\infty$	likely linear
$\ln \tilde{X}_5$	5288.754	$+\infty$	likely linear
$\ln \tilde{X}_6$	7059.638	$+\infty$	likely linear
$\ln Y_1$	0.443	$+\infty$	nonlinear
$\ln Y_2$	0.597	$+\infty$	nonlinear
$\ln Y_3$	0.103	$+\infty$	nonlinear
$\ln Y_4$	2.511	$+\infty$	nonlinear
t	3.002	$+\infty$	nonlinear
Categorical			
id	0.006	1	relevant

Bandwidths are selected via least-squares cross-validation using local linear least squares.

The gradient estimates for the unconstrained nonparametric estimator are given in Table 4. Similar to the parametric estimates, we do not see violations in the quartile values. Although there appear to be more violations than in the parametric model, the fit of the model is slightly higher (0.9973).

To get a better feel for the estimates, we plot their kernel density estimates in Figure 4. There are at least two points worth mentioning with regard to the unconstrained estimates (dashed lines). First, there is substantially more variation in the estimates than we saw with the parametric results, a common finding when comparing parametric and nonparametric models given the lack of structure placed *a priori* on the problem by nonparameric methods. These differences are likely exacerbated by the bandwidth on *id*. Second, there appears to be a larger percentage of observations for which the estimated gradients have the wrong sign. In fact, it appears that each of the gradients have violations. In fact, Table 4 shows that the percentage of violations from the nonparametric model are generally greater than for the parametric model. This example shows that if we wish

to use nonparametric methods for the estimation of distance functions, we likely need to employ a constrained estimator.

TABLE 4. **Summary Statistics of Partial Derivatives: Nonparametric**

	$\frac{\partial \ln X_1}{\partial \ln X_2}$	$\frac{\partial \ln X_1}{\partial \ln X_3}$	$\frac{\partial \ln X_1}{\partial \ln X_4}$	$\frac{\partial \ln X_1}{\partial \ln X_5}$	$\frac{\partial \ln X_1}{\partial \ln X_6}$	$\frac{\partial \ln X_1}{\partial \ln Y_1}$	$\frac{\partial \ln X_1}{\partial \ln Y_2}$	$\frac{\partial \ln X_1}{\partial \ln Y_3}$	$\frac{\partial \ln X_1}{\partial \ln Y_4}$
Unrestricted									
Min	-0.5812	-0.6271	-0.5065	-0.6737	-0.3296	-0.2155	-0.0841	-0.4989	-0.1017
	(0.1004)	(0.1369)	(0.0968)	(0.0624)	(0.0642)	(0.0174)	(0.0149)	(0.0244)	(0.0336)
Q1	-0.1921	-0.2312	-0.1364	-0.3628	-0.0689	0.3422	0.1572	0.2556	0.0043
	(0.0152)	(0.0250)	(0.0066)	(0.0237)	(0.0046)	(0.0005)	(0.0006)	(0.0025)	(0.0001)
Median	-0.1500	-0.1963	-0.0947	-0.3101	-0.0505	0.3974	0.2055	0.4070	0.0128
	(0.0009)	(0.0010)	(0.0008)	(0.0011)	(0.0004)	(0.0011)	(0.0008)	(0.0033)	(0.0002)
Mean	-0.1538	-0.1913	-0.0959	-0.3077	-0.0502	0.3884	0.1979	0.4043	0.0131
	(0.0026)	(0.0032)	(0.0018)	(0.0048)	(0.0010)	(0.0061)	(0.0032)	(0.0073)	(0.0003)
Q3	-0.1141	-0.1556	-0.0551	-0.2551	-0.0307	0.4424	0.2477	0.5518	0.0214
	(0.0006)	(0.0005)	(0.0006)	(0.0008)	(0.0003)	(0.0282)	(0.0267)	(0.0392)	(0.0006)
Max	0.1565	0.1257	0.3446	0.3944	0.1697	0.7532	0.5182	1.3801	0.1026
	(0.0156)	(0.0402)	(0.0813)	(0.1020)	(0.0321)	(0.0675)	(0.0539)	(0.0995)	(0.0173)
Pct. Of Violation	1.80%	1.20%	6.97%	0.39%	6.72%	0.37%	1.71%	3.69%	15.62%
Restricted									
Min	-0.5300	-0.6275	-0.4052	-0.5748	-0.2299	0.0000	0.0000	0.0000	0.0000
	(0.0918)	(0.1465)	(0.0382)	(0.0172)	(0.0259)	(0.0425)	(0.0131)	(0.0038)	(0.0082)
Q1	-0.1890	-0.2232	-0.1454	-0.3548	-0.0761	0.3404	0.1576	0.2662	0.0098
	(0.0159)	(0.0212)	(0.0123)	(0.0239)	(0.0072)	(0.0006)	(0.0005)	(0.0021)	(0.0001)
Median	-0.1495	-0.1895	-0.1058	-0.3052	-0.0591	0.3924	0.2047	0.4112	0.0175
	(0.0007)	(0.0010)	(0.0007)	(0.0011)	(0.0004)	(0.0011)	(0.0010)	(0.0034)	(0.0002)
Mean	-0.1526	-0.1850	-0.1099	-0.3019	-0.0597	0.3843	0.1986	0.4131	0.0187
	(0.0026)	(0.0029)	(0.0019)	(0.0047)	(0.0010)	(0.0059)	(0.0033)	(0.0074)	(0.0003)
Q3	-0.1146	-0.1505	-0.0690	-0.2500	-0.0408	0.4352	0.2457	0.5501	0.0252
	(0.0005)	(0.0005)	(0.0006)	(0.0007)	(0.0002)	(0.0280)	(0.0245)	(0.0476)	(0.0020)
Max	0.0000	0.0000	0.0000	0.0000	0.0000	0.8556	0.5031	1.1353	0.1262
	(0.0315)	(0.0518)	(0.0071)	(0.0017)	(0.0053)	(0.1556)	(0.0440)	(0.0516)	(0.0410)
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

1. The numbers in the parentheses are standard errors. Standard errors are calculated using a wild bootstrap with 99 replications.
2. The *p*-value is the probability of rejecting the null that the constraint is valid when it is true.

Although we find larger percentages of violations, the nonparametric model is not without its merits. In addition to a larger goodness-of-fit, the nonparametric model produces more reasonable estimates of median RTS (0.9802 with standard error 0.0020).[18] Even with the violations included,

---

[18]We will consider more formal comparisons in Section 5.3.

the density has a median value near unity which is more consistent with expectations. That being said, there is non-trivial variation in the estimates.

FIGURE 4. **Gradient estimates from the unconstrained and constrained Nonparametric model**

---- Unconstrained ⎯⎯ Constrained
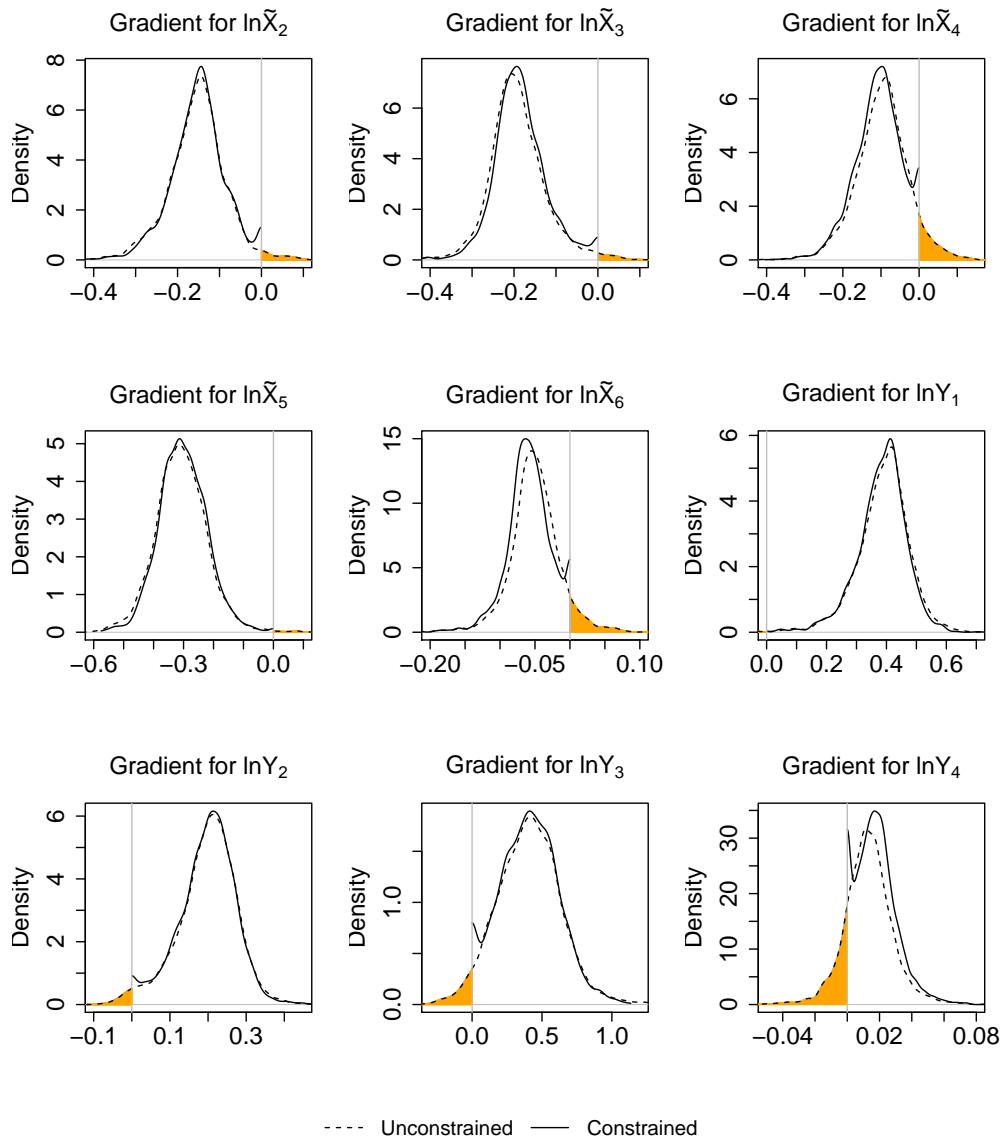
5.2.2. *Constrained Estimator.* In comparison to the parametric model, it may be more necessary in the nonparametric model to impose constraints given the additional violations present in our application. Therefore we employ the CWB procedure on the nonparametric estimates. We use

the bandwidths calculated by the unconstrained model and then employ our constrained estimator using those bandwidths. The summary of the gradients for the constrained nonparametric estimator are also given in Table 4. By construction, each of these values takes the appropriate sign and are significant. Again, we do not see a uniform trend in terms of the magnitude change in the gradients.

In order to get the full picture for the gradient estimates, Figure 4 plots the kernel density functions of the constrained gradient estimates (solid lines). The densities for the constrained gradients are again plotted using the Silverman reflection method. We see that with the constrained model we have masses near zero, especially for the gradients of $\ln \tilde{X}_4$, $\ln \tilde{X}_6$, $\ln Y_3$ and $\ln Y_4$. One other difference is with respect to the estimates which did not violate economic theory. Here the constrained and unconstrained estimators do not differ as much. In short, we see a lot of the negative estimates moved just to the other side of zero and relatively small changes in the other gradient values (say, from $\ln Y_2$ to $\ln Y_4$). This difference occurs because the nonparametric estimator is a local as opposed to a (parametric) global estimator. The local estimator allows us to only modify the function in the area where the assumptions are violated and leaves the other results virtually unchanged. This is exactly what we saw in the graphical illustrations.

The non-negligible bumps at the boundaries of the constrained kernel density estimates could potentially represent an underlying feature of the data generating process with respect to the theoretical constraints (for example the density of the gradient for $\ln Y_4$). However, for many of the smaller bumps, we feel that these are nothing more than artifacts of the reflection method and the corresponding bandwidth (for example the density of the gradient for $\ln Y_2$). Notice that the largest 'spikes' at zero are for the constrained gradient estimates from the nonparametric model with the highest percentage of violations. We also mention that the violations and subsequent spikes at zero are inherently linked to the level of smoothing deployed prior to enforcing the constraints.

It was argued previously that the negative gradients from the unconstrained estimator may falsely provide intuitive estimates of RTS. However, even though there is a slight shift to the right in the density, the median RTS is 0.9759 with standard error 0.0020, slightly lower than the unconstrained counterpart. The constrained estimates did not affect this intuitive result.

In summary, the unconstrained nonparametric estimator brought increased percentages of violations of economic theory. This likely resulted from the increased level of flexibility as well as

the bandwidth on the farm id variable. However, the results for RTS were more in line with the popular view and the in-sample fit is higher for the nonparametric model. Finally, we saw that the nonparametric model (effectively) only changed the function where it was violated, thus leaving the "well-behaved" points as they were.

5.3. **Comparison of Methodologies.** The previous sub-sections did not produce a clear domination by any particular estimator: parametric or nonparametric, unconstrained or constrained. Therefore we will consider formal tests as well as out-of-sample predictive performance. We first employ the Hsiao, Li & Racine (2007) test for correct parametric specification and then perform out-of-sample predictions by taking a subset of the data out of the sample and then try to predict on it using the remaining data.

5.3.1. *Specification.* To test for correct parametric specification, we employ the test by Hsiao et al. (2007) which allows for mixed categorical and continuous data. The null hypothesis is that the unconstrained parametric model is correctly specified ($H_0 : \Pr[E(y|x) = g(x; \beta)] = 1$) against the alternative that it is not ($H_1 : \Pr[E(y|x) = g(x; \beta)] < 1$). The test statistic is based on $I_n \equiv E\left(E\left(\varepsilon|x\right)^2 f(x)\right)$, where $\varepsilon = y - g(x; \beta)$ represents the residuals from the parametric model under investigation and $n \equiv \sum_{i=1}^{N} T_i$ is the sample size, where N is the number of cross-sectional units, and T_i is the number of time periods for the i-th cross-sectional unit. I_n is non-negative and equals zero if and only if the null is true. The standardized test statistic $\hat{J}_n = n(\hat{\gamma}_{c_1} \dots \hat{\gamma}_{c_q})^{1/2} \hat{I}_n / \sqrt{\hat{\sigma}_J} \to \mathcal{N}(0, 1)$ in distribution under the null; $\hat{\gamma}_{c_1} \dots \hat{\gamma}_{c_q}$ denote bandwidths of continuous regressors; The statistic $\hat{I}_n = n^{-1} \sum_{i=1}^{n} \hat{u}_i \hat{E}_{-i}(u_i \mid x_i) \hat{f}_{-i}(x_i)$ was independently proposed by Fan & Li (1996) and Zheng (1996), where $\hat{u}_i$ is the parametric null model's residual and $\hat{E}_{-i}(u_i \mid x_i) \hat{f}_{-i}(x_i)$ is a leave-one-out kernel estimator. Finally, $\hat{\sigma}_J^2 = \frac{2(\hat{\gamma}_1 \dots \hat{\gamma}_{q_c})}{n^2} \sum_i \sum_{j \neq i} \hat{u}_i^2 \hat{u}_j^2 K_\gamma^2(X_i, x_j)$. Hsiao et al. (2007) advocate the residual-based wild bootstrap method to approximate the null distribution of $\hat{J}_n$. From the wild bootstrap residual $\{u_i^*\}_{i=1}^{n}$, we generate $y_i^* = g(x_i, \widehat{\beta}) + u_i^*$. We then use the bootstrap sample $\{y_i^*, x_i\}$ to obtain a least-squares estimator of β. The bootstrap residual is given by $\widehat{u}_i^* = y_i^* - g(x_i, \hat{\beta}^*)$. The bootstrap test statistic $\hat{J}_n^*$ is obtained from the same formula for $\hat{J}_n$, with $\hat{u}_i$ being replaced by $\hat{u}_i^*$. We repeat the above steps $B = 399$ times, the original test statistic $\hat{J}_n$ plus the B bootstrap test statistics give us the empirical distribution of the bootstrap statistics, which is then used to approximate the finite-sample null distribution of $\hat{J}_n$. Using the

wild bootstrap gives a p-value that is zero to four decimal places. In other words, we reject the null that the unconstrained translog model is correctly specified.[19]

5.3.2. *Prediction.* Even though we reject the parametric model and find that the nonparametric model has a better in-sample fit, it is not necessarily the case that the nonparametric model results in better out-of-sample prediction. Nonparametric models are often criticized for overfitting the data.

In order to compare the prediction performance of the two estimators we present the results from the following exercise. Recall that we have a unbalanced panel data with $N = 701$ farms and each farm is measured for T_i years. We chose the first $T_i - 1$ years of data for each farm as our training data. We then forecast over the remaining year. For the translog model, the predicted MSE for the unconstrained model was 0.00626 and this number was 0.00625 for the constrained model. In other words, for the translog model, the out-of-sample prediction is (slightly) better for the constrained model. That being said, the difference is only seen at the fifth decimal place. For the nonparametric model, the predicted MSE for the unconstrained and constrained estimators are 0.0065 and 0.0052, respectively. Here we see improvement in terms of out-of-sample prediction for the nonparametric model (on the magnitude of 25%). This suggests that the unconstrained estimator may be modeling noise and the constrained estimator is better at modeling the true underlying process.

In terms of the comparison between the nonparametric and parametric models, the parametric model does a slightly better job at out-of-sample prediction for the unconstrained model (i.e., the predicted MSE are 0.0063 vs. 0.0065). However, the nonparametric model does a better job for the constrained model (i.e., the predicted MSE are 0.0052 vs. 0.0063). We note here that this outcome is not necessarily true for all nonparametric models as there are examples of both better (Henderson & Kumbhakar 2006) and worse (Henderson & Millimet 2008) out-of-sample prediction as compared to parametric models in the existing literature.

5.4. **Testing.** Looking at the percentage of violations is relevant, but it is important to formally test for the presence of violations. We present a method to test for the null hypothesis of monotonicity

---

[19]We are also interested in testing the null that the constrained translog model is correctly specified. Unfortunately, it is unclear how to perform the HLR test in this scenario. We expect that this model will also be rejected, as the deviations from the unconstrained model are minimal, but we do not make a formal claim at this time.

in our gradient estimates. Using a similar methodology, we also consider a test for constant returns to scale (CRS). Although it is not required by theory in this application, it is interesting to see if a model that imposes CRS would be appropriate. This requires us to constrain the sum of four of our gradients to be equal to unity. We report the monotonicity test results first.

5.4.1. *Monotonicity.* As mentioned before, we constrain the gradients of nine regressors simultaneously. Specifically, for each observation, we require $\frac{\partial \ln X_1}{\partial \ln X_q} \leq 0$ for $q = 2, \ldots, 6$, and $\frac{\partial \ln X_1}{\partial \ln Y_m} \geq 0$ for $m = 1, \ldots, 4$.

We again note that we have nine joint constraints to impose over 4333 observations. This is rejected at the 1% level for the nonparametric model. This is not surprising given there are violations for each variable in Table 4. It is also not surprising that this null is also rejected for the translog model at the 1% level as there were also gradients with substantial violations.

We also tested each of these constraints individually. Tables 2 and 4 report the *p*-values of testing if the individual restriction is valid. It turns out, in the nonparametric model, that we reject the null hypothesis of each constraint holding everywhere at the 1% level individually. This is perhaps surprising given that some of the gradients had very few violations. These results can be reconciled with intuition. The constrained model imposes the monotonicity constraint everywhere and the findings of the test could be driven by only a few points where the gradients are substantially negative, leading to a dramatic difference between the uniform weights and the constraint weights. Even if there is a small number of violations (as is the case with the parametric translog model), the violations could require the perturbation of the uniform weights such that the test rejects the null hypothesis. It is not that a larger percentage of violations must produce a smaller *p*-value, it is how the weights have to change to ensure that the constraints are violated. Notice from Table 4 when the gradient estimates are of the wrong sign they are quite distant from zero, the binding case. Thus, our rejection of monotonicity of the estimated IDF in each dimension is more likely a product of the common symptoms underlying most economic data rather than an indictment on the underlying theory.

Contrast the rejection of each individual constraint with the results from the translog model in which we have six cases where we fail to reject the null of monotonicity at the 1% level. It should come as no surprise that one of them is for the monotonicity constraint on the gradients

of $\ln \tilde{X}_5$ since we did not have a single violation in the translog model. We also fail to reject the monotonicity in $\ln \tilde{X}_2$, $\ln \tilde{X}_4$, $\ln \tilde{X}_6$, $\ln Y_1$, and $\ln Y_3$, individually, at the 1% level. We do, however, reject the monotonicity in $\ln \tilde{X}_3$, $\ln Y_2$, and $\ln Y_4$ at the 1% level, with zero (to four decimal places) p-values. The likely insensitivity of the individual constraints to minor violations in the translog model relative to the nonparametric model stems from the fact that this estimator is global and the constraint weights are less influenced by a few deviant gradients than the nonparametric method.

5.4.2. *Constant Returns to Scale.* A separate constraint of interest in many economic production models is the assumption of constant returns to scale (CRS). This is a common assumption in many models and we seek to test it for our particular application. This constraint is different from the constraints imposed previously as we look to a combination of gradient vectors. Specifically, CRS holds in our model if $\sum_{m=1}^{4} \partial \ln X_1 / \partial \ln Y_m = 1$. We note that while this is different in principle from the other constraints, it requires minimal additional effort in terms of computational need. These partial derivatives are still linear in p meaning that they can easily be incorporated with our previous constraints.

The results for the parametric models are as expected. The null of CRS is rejected at the 1% level for both the unconstrained and constrained models. The same null is also rejected for the nonparametric models. Although the median estimates were near unity, there was enough variation in the RTS estimates and hence we reject the null. Recall that CRS is generally rejected for a technology with more than two inputs and the null is that CRS holds for all observations. Hence, at least some of the observations have non-unity returns to scale. Here we have evidence of why it may be naïve to solely examine the data at a mean value.

6. Conclusion

Estimating regression models consistent with economic theory are crucial for providing sound assessment of economic phenomena. Methods to guarantee that the conditions set forth in theoretical models often are limited to a particular style of estimation or involve rigorous computational devices to impose the constraints of interest. In this paper we set forth a general methodology for the imposition of smoothness constraints on economic relationships using linear regression estimators with a technique known as constraint weighted bootstrapping. While known in the nonparametric

regression literature, this methodology can also be used in parametric regression. To investigate the validity of the constraints, we propose a test of the smoothness constraints imposed that allows for heteroscedasticity and is easily implementable.

We employ these methods to study a multi-input, multi-output unbalanced panel data set of Norwegian dairy farms. We found violations of economic constraints in both our parametric and nonparametric estimators and employed the techniques in the paper to correct for these violations. The nonparametric model, not surprisingly, had more violations of monotonicity that the parametric model, however, the parametric model had an uncharacteristically high assessment of returns to scale over the nonparametric model, especially given that both models fit the data well.

While the unconstrained nonparametric model produced numerous sign violations of the estimated gradients, the constrained methods did not alter the data dramatically enough to change the overarching insights from the distance function. This is comforting given that the constrained methods could have overly restricted the flexibility of the nonparametric model to ensure the sign constraints were satisfied, but this did not appear to be the case. Given that the fit of the constrained nonparametric model was not adversely impacted from imposition of the constraints, this lends further credence to these methods. When comparing across the two methods we found that we were able to reject the parametric functional form and that we had improved out-of-sample prediction with the nonparametric model, especially when employing the constrained estimator.

REFERENCES

Abrevaya, J. & Jiang, W. (2005), 'A nonparametric approach to measuring and testing curvature', *Journal of Business and Economic Statistics* **23**, 1–19.

Ajgaonkar, S. G. P. (1965), 'On a class of linear estimators in sampling with varying probabilities without replacement', *Journal of the American Statistical Association* **60**(310), 637–642.

Cressie, N. A. C. & Read, T. R. C. (1984), 'Multinomial goodness-of-fit tests', *Journal of the Royal Statistical Society, Series B* **46**, 440–464.

Diewert, W. (1992), 'Fisher ideal output, input, and productivity indexes revisited', *Journal of Productivity Analysis* **3**, 211–248.

Diewert, W. E. & Wales, T. J. (1987), 'Flexible functional forms and global curvature conditions', *Econometrica* **55**(1), 43–68.

Du, P., Parmeter, C. F. & Racine, J. S. (2012), 'Nonparametric kernel regression with multiple predictors and multiple shape constraints', *Statistica Sinica* . forthcoming.

Epstein, L. G. & Yatchew, A. J. (1985), 'Nonparametric hypothesis testing procedures and applications to demand analysis', *Journal of Econometrics* **30**, 149–169.

Fan, Y. & Li, Q. (1996), 'Consistent model specification tests: Omitted variables and semiparametric functional forms', *Econometrica* **64**, 865–890.

Färe, R. & Primont, D. (1995), *Multi-Output Production and Duality: Theory and Applications*, Kluwer.

Hall, P. & Huang, H. (2001), 'Nonparametric kernel regression subject to monotonicity constraints', *The Annals of Statistics* **29**(3), 624–647.

Hall, P., Huang, H., Gifford, J. & Gijbels, I. (2001), 'Nonparametric estimation of hazard rate under the constraint of monotonicity', *Journal of Computational and Graphical Statistics* **10**(3), 592–614.

Härdle, W. & Mammen, E. (1993), 'Comparing nonparametric versus parametric regression fits', *Annals of Statistics* **21**(4), 1926–1947.

Henderson, D. J. & Kumbhakar, S. C. (2006), 'Public and private capital productivity puzzle: A nonparametric approach', *Southern Economic Journal* **73**(2), 219–232.

Henderson, D. J. & Millimet, D. L. (2008), 'Is gravity linear?', *Journal of Applied Econometrics* **23**(1), 137–172.

Henderson, D. J. & Parmeter, C. F. (2009), Imposing economic constraints in nonparametric regression: Survey, implementation and extension, *in* Q. Li & J. S. Racine, eds, 'Advances in Econometrics: Nonparametric Econometric Methods', Vol. 25, Emerald Group Publishing.

Hsiao, C., Li, Q. & Racine, J. S. (2007), 'A consistent model specification test with mixed categorical and continuous data', *Journal of Econometrics* **140**, 802–826.

Hurvich, C. M., Simonoff, J. S. & Tsai, C. L. (1998), 'Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion', *Journal of the Royal Statistical Society Series B* **60**, 271–293.

Kumbhakar, S. C. (2012), 'Estimation of production technology using primal approaches', *European Journal of Operational Research* **217**(3), 509–518.

Kumbhakar, S. C. & Lovell, C. A. K. (2000), *Stochastic Frontier Analysis*, Cambrdige University Press, New York, NY.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Matzkin, R. L. (1994), Restrictions of economic theory in nonparametric methods, *in* D. L. McFadden & R. F. Engle, eds, 'Handbook of Econometrics', Vol. 4, North-Holland: Amsterdam.

McFadden, D. L. (1978), Cost, revenue and profit functions, *in* M. Fuss & D. L. McFadden, eds, 'Production Economics: A Dual Approach to Theory and Applications', Vol. 1: The Theory of Production, pp. 2–109.

O'Donnell, C. J. & Coelli, T. J. (2005), 'A Bayesian approach to imposing curvature on distance functions', *Journal of Econometrics* **126**(2), 493–523.

O'Donnell, C. J., Rambaldi, A. N. & Doran, H. E. (2001), 'Estimating economic relationships subject to firm- and time-varying equality and inequality constraints', *Journal of Applied Econometrics* **16**(4), 709–726.

Parmeter, C. F. & Racine, J. S. (2012), Smooth constrained frontier analysis, *in* X. Chen & N. E. Swanson, eds, 'A Festschrift in Honour of Halbert L. White Jr.', Springer-Verlag.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Racine, J. S. & Li, Q. (2004), 'Nonparametric estimation of regression functions with both categorical and continuous data', *Journal of Econometrics* **119**(1), 99–130.

Rambaldi, A. N. & Doran, H. E. (1997), 'Applying linear time-varying constraints to econometric models: with an application to demand systems', *Journal of Econometrics* **79**(1), 83–95.

Ryan, D. L. & Wales, T. J. (2000), 'Imposing local concavity in the translog and generalized Leontief cost functions', *Economics Letters* **67**, 253–260.

Sauer, J. (2006), 'Economic theory and econometric practice: parametric efficiency analysis', *Empirical Economics* **31**, 1061–1087.

Shephard, R. W. (1953), *Cost and production functions*, Princeton University Press.

Shephard, R. W. (1970), *The theory of cost and production functions*, Princeton University Press.

Terrell, D. (1996), 'Incorporating monotonicity and concavity conditions in flexible functional forms', *Journal of Applied Econometrics* **11**(2), 179–194.

Yatchew, A. & Bos, L. (1997), 'Nonparametric regression and testing in economic models', *Journal of Quantitative Economics* **13**, 81–131.

Zheng, J. (1996), 'A consistent test of functional form via nonparametric estimation technique', *Journal of Econometrics* **75**, 263–289.

A. Translog Parameter Estimates

The parameter estimates for the unconstrained translog model with fixed effects are given in Table 5. Excluding the farm effects, we first note that 33 out of the possible 65 parameter estimates (excluding intercepts) are significant at the 10% level. We also note that the fit of the model is very high with a squared correlation coefficient between the actual and fitted values equal to 0.9845.

TABLE 5. **Translog with State Effects: Unrestricted vs. Restricted**

	Unrestricted		Restricted	
	Estimate	Standard error	Estimate	Standard error
α_0	2.8610	(0.9167)	2.5793	(0.9152)
β_2	0.1347	(0.1695)	-0.0961	(0.1693)
β_3	-0.1817	(0.1446)	-0.1905	(0.1443)
β_4	0.5333	(0.1868)	0.3109	(0.1865)
β_5	-0.8259	(0.2111)	-0.4837	(0.2108)
β_6	-0.1238	(0.1139)	-0.1110	(0.1137)
α_1	0.4493	(0.1902)	0.5109	(0.1899)
α_2	-0.1878	(0.1165)	-0.4050	(0.1163)
α_3	0.1362	(0.3712)	0.2086	(0.3705)
α_4	-0.0395	(0.0331)	-0.0233	(0.0331)
α_t	-0.0301	(0.0135)	-0.0236	(0.0135)
β_{22}	-0.0125	(0.0292)	0.0057	(0.0292)
β_{23}	0.0107	(0.0170)	-0.0072	(0.0170)
β_{24}	0.0782	(0.0216)	0.0487	(0.0216)
β_{25}	-0.0635	(0.0248)	-0.0493	(0.0247)
β_{26}	-0.0133	(0.0128)	-0.0113	(0.0128)
δ_{21}	0.1387	(0.0231)	0.1398	(0.0230)
δ_{22}	-0.0384	(0.0134)	-0.0383	(0.0134)
δ_{23}	-0.1337	(0.0362)	-0.0813	(0.0361)
δ_{24}	0.0113	(0.0038)	0.0041	(0.0038)
β_{2t}	-0.0062	(0.0015)	-0.0048	(0.0015)
β_{33}	-0.0788	(0.0076)	-0.0348	(0.0076)
β_{34}	-0.0082	(0.0157)	0.0216	(0.0156)
β_{35}	0.0637	(0.0210)	0.0206	(0.0210)
β_{36}	-0.0112	(0.0100)	-0.0115	(0.0100)
δ_{31}	-0.0999	(0.0184)	-0.0597	(0.0184)
δ_{32}	0.0285	(0.0121)	0.0047	(0.0120)
δ_{33}	0.0626	(0.0289)	0.0461	(0.0288)
δ_{34}	-0.0026	(0.0033)	0.0007	(0.0033)
β_{3t}	0.0026	(0.0013)	0.0010	(0.0012)
β_{44}	-0.0592	(0.0324)	-0.0710	(0.0323)
β_{45}	0.0068	(0.0250)	0.0103	(0.0250)
β_{46}	-0.0143	(0.0129)	-0.0034	(0.0129)
R^2	0.9845		0.9845	
δ_{41}	0.0171	(0.0242)	-0.0057	(0.0242)
δ_{42}	0.0062	(0.0136)	-0.0062	(0.0136)
δ_{43}	-0.0696	(0.0434)	-0.0267	(0.0434)
δ_{44}	-0.0089	(0.0042)	-0.0019	(0.0042)
β_{4t}	-0.0006	(0.0016)	-0.0005	(0.0016)
β_{55}	0.0259	(0.0388)	0.0196	(0.0387)
β_{56}	0.0025	(0.0146)	0.0070	(0.0146)
δ_{51}	-0.0611	(0.0266)	-0.0743	(0.0265)
δ_{52}	-0.0399	(0.0163)	-0.0065	(0.0162)
δ_{53}	0.1415	(0.0465)	0.0696	(0.0464)
δ_{54}	0.0014	(0.0043)	-0.0010	(0.0043)
β_{5t}	0.0002	(0.0018)	-0.0007	(0.0018)
β_{66}	0.0055	(0.0115)	-0.0011	(0.0115)
δ_{61}	0.0184	(0.0144)	0.0035	(0.0143)
δ_{62}	0.0104	(0.0086)	0.0145	(0.0085)
δ_{63}	-0.0215	(0.0246)	-0.0163	(0.0245)
δ_{64}	-0.0094	(0.0023)	-0.0032	(0.0023)
β_{6t}	0.0030	(0.0010)	0.0032	(0.0010)
α_{11}	0.3943	(0.0362)	0.2627	(0.0361)
α_{12}	-0.0433	(0.0154)	-0.0013	(0.0153)
α_{13}	-0.2024	(0.0418)	-0.1406	(0.0417)
α_{14}	-0.0015	(0.0043)	-0.0021	(0.0043)
α_{1t}	-0.0093	(0.0017)	-0.0072	(0.0017)
α_{22}	0.0575	(0.0072)	0.0144	(0.0072)
α_{23}	0.0148	(0.0267)	0.0515	(0.0267)
α_{24}	-0.0002	(0.0024)	0.0000	(0.0024)
α_{2t}	-0.0014	(0.0011)	-0.0005	(0.0011)
α_{33}	0.0519	(0.0882)	-0.0005	(0.0881)
α_{34}	0.0176	(0.0073)	0.0103	(0.0073)
α_{3t}	0.0114	(0.0030)	0.0088	(0.0030)
α_{44}	0.0051	(0.0008)	0.0012	(0.0008)
α_{4t}	0.0007	(0.0003)	0.0005	(0.0003)
α_{tt}	-0.0016	(0.0002)	-0.0016	(0.0002)
R^2	0.9845		0.9845	

1. The numbers in the parentheses are standard errors.
2. The estimates of fixed effects are omitted to save space, but are available upon request.