Shape-Constrained Kernel-Weighted Least Squares: Estimating Production Functions for Chilean Manufacturing Industries

Daisuke YAGI
Texas A&M University, College Station, TX 77843 (*d.yagi@tamu.edu*)

Yining CHEN
London School of Economics and Political Science, London, WC2A 2AE, UK (*y.chen101@lse.ac.uk*)

Andrew L. JOHNSON
Texas A&M University, College Station, TX 77843; Osaka University, Osaka, Japan (*ajohnson@tamu.edu*)

Timo KUOSMANEN
Aalto University School of Business, Helsinki, Finland (*Timo.Kuosmanen@aalto.fi*)

In this article, we examine a novel way of imposing shape constraints on a local polynomial kernel estimator. The proposed approach is referred to as shape constrained kernel-weighted least squares (SCKLS). We prove uniform consistency of the SCKLS estimator with monotonicity and convexity/concavity constraints and establish its convergence rate. In addition, we propose a test to validate whether shape constraints are correctly specified. The competitiveness of SCKLS is shown in a comprehensive simulation study. Finally, we analyze Chilean manufacturing data using the SCKLS estimator and quantify production in the plastics and wood industries. The results show that exporting firms have significantly higher productivity.

KEY WORDS: Kernel estimation; Local polynomials; Multivariate convex regression; Nonparametric regression; Shape constraints

1. INTRODUCTION

Nonparametric regression methods, such as the local linear (LL) estimator, avoid functional form misspecification. To model production with a production or a cost function, the flexible nature of nonparametric methods can cause difficulties in interpreting the results. Fortunately, microeconomic theory provides additional structure in the form of shape constraints. Recently, several nonparametric shape constrained estimators have been proposed that combine the advantage of avoiding parametric functional specification with improved small-sample performance relative to unconstrained nonparametric estimators. Nevertheless, the existing methods have limitations regarding either estimation performance or computational feasibility. In this article, we propose a new estimator that imposes shape restrictions on local kernel weighting methods. By combining local averaging with shape constrained estimation, we improve finite sample performance by avoiding overfitting.

Work on shape-constrained regression first started in the 1950s with Hildreth (1954), who studied the univariate regressor case with a least-squares objective subject to monotonicity and concavity/convexity constraints. See also Brunk (1955) and Grenander (1956) for alternative shape constrained estimators. Under the concavity/convexity constraint, properties such as consistency, rate of convergence, and asymptotic distribution have been shown by Hanson and Pledger (1976), Mammen (1991), and Groeneboom, Jongbloed, and Wellner (2001), respectively. In the multivariate case, Kuosmanen (2008) developed the characterization of the least-squares estimator subject to concavity/convexity and monotonicity constraints, which we will refer to as convex nonparametric least squares (CNLS) throughout this article. Furthermore, consistency of the least-squares estimator was shown independently by Seijo and Sen (2011) and Lim and Glynn (2012).

Regarding the nonparametric estimation implemented using kernel based methods, Birke and Dette (2007), Carroll, Delaigle, and Hall (2011), and Hall and Huang (2001) investigated the univariate case and proposed smooth estimators that can impose derivative-based constraints including monotonicity and concavity/convexity. Du, Parmeter, and Racine (2013) proposed constrained weighted bootstrap (CWB) by generalizing Hall and Huang's method to the multivariate regression setting. Beresteanu (2007) developed a similar type of estimator but for use with spline-based estimators. Finally, we mention the work of Li, Liu, and Li (2017), which extended Hall and Huang's method to use the k-nearest neighbor approach subject to the monotonicity constraint.

Journal of Business & Economic Statistics
2018, Vol. 0, No. 0
DOI: 10.1080/07350015.2018.1431128

In this article, *shape constrained kernel-weighted least squares* (SCKLS) estimator is described, which optimizes a local polynomial kernel criterion while estimating a multivariate regression function with shape constraints. Under the monotonicity and convex/concavity constraints, we prove uniform consistency and establish the convergence rate of the SCKLS estimator. Kuosmanen (2008), Seijo and Sen (2011), and Lim and Glynn (2012) emphasized the potential advantage that CNLS does not require the selection of tuning parameters. Our proposed SCKLS estimator sheds further light on this issue: in the SCKLS framework, CNLS can be seen as the zero bandwidth estimator; we argue that, compared to unrestricted kernel methods, the SCKLS estimator is relatively robust to the bandwidth selected and is able to alleviate well-known issues such as boundary inconsistency faced by the CNLS estimator.

Note that with n observations, CNLS imposes $O(n^2)$ concavity/convexity constraints, which can lead to computational difficulties. The number of constraints and the number of variables in the SCKLS estimator do not depend on the number of observations, but rather the number of evaluation points which is arbitrarily defined by the modeler, thereby bring the computational complexity of the estimator largely under control of the modeler. In this article, we implement an iterative algorithm that reduces the number of constraints by building on the ideas in Lee et al. (2013) to further improve the computational performance. We then validate the performance of the SCKLS estimator via Monte Carlo simulations. For a variety of parameter settings, we find performance of SCKLS to be better or at least competitive with CNLS, CWB, and the local linear estimators. We provide the first simulation study of CWB with global concavity constraints. We also investigate the use of variable bandwidth methods that are a function of the data density[1] and propose variants of a uniform grid as practical ways to further improve the performance of SCKLS.

Crucially, we also investigate the behavior of SCKLS when the shape constraints are misspecified and propose a hypothesis test to validate the shape constraints imposed. Having a test that validates the shape constraints is critical because otherwise our estimation procedure would lead to inconsistent estimates.

Finally, we apply the SCKLS estimator empirically on Chilean manufacturing data from the Chilean Annual Industrial Survey. The estimation results provide a concise description of the supply-side of the Chilean plastic and wood industries as we report marginal productivity, marginal rate of substitution, and most productive scale size. We also investigate the impact of exporting on productivity by including additional predictors of output in a semi-parametric model. We find that exporting correlates with higher productivity, thus supporting international trade theories that high productivity firms are more likely to compete in international markets.

Our focus on production functions guides our selection of the polynomial function used in estimation, the data-generation processes (DGP) in the Monte Carlo simulations. For the application analyzing the Chilean manufacturing data, we are interested in monotonic and concave shape constraints and use a local linear kernel function. These assumptions are motivated by standard economic theory for production functions (Varian 1984). However, the methods proposed in the article are general and applicable for other applications with higher order polynomial functions or alternative shape restrictions, as discussed in Appendix A.

The remainder of this article is as follows. Section 2 describes the model framework and presents our estimator, SCKLS. Section 3 contains the statistical properties of the estimator, and Section 4 discusses the behavior of SCKLS under misspecification, as well as a test for concavity and monotonicity. Monte Carlo simulation results under several different experimental settings are shown in Section 5. Section 6 applies the SCKLS estimator to estimate a production function for both the Chilean plastics and wood industries. Section 7 concludes and suggests future research directions. Appendix A provides extensions to SCKLS and a comparison to CNLS and CWB. Appendix B contains all the technical proofs and Appendix C describes a test for affinity. Appendix D states the details of the iterative algorithm for SCKLS, and Appendix E presents a more extensive set of simulation results. Appendix F describes the details of the partially linear model, and Appendix G gives further details about the application to the Chilean manufacturing data. All appendices are available as supplemental materials via the *Journal of Business & Economic Statistics* website.

2. MODEL FRAMEWORK AND METHODOLOGY

2.1 Model

Suppose we observe n pairs of input and output data, $\{\boldsymbol{X}_j, y_j\}_{j=1}^n$, where for every $j = 1, \ldots, n$, $\boldsymbol{X}_j = (X_{j1}, \ldots, X_{jd})' \in \mathbb{R}^d$ is a d-dimensional input vector, and $y_j \in \mathbb{R}$ is an output. Consider the following regression model

$$y_j = g_0(\boldsymbol{X}_j) + \epsilon_j, \qquad \text{for } j = 1, \ldots, n,$$

where ϵ_j is a random variable satisfying $E(\epsilon_j|\boldsymbol{X}_j) = 0$. Assume that the regression function $g_0 : \mathbb{R}^d \to \mathbb{R}$ belongs to a class of functions, G, that satisfies certain shape restrictions. Here, our estimator can impose any shape restriction that can be modeled as a lower or upper bound on a derivative. Examples are supermodularity, convexity, monotonicity, and quasi-convexity. For purposes of concreteness, and in view of the application to production functions, we focus on imposing monotonicity and global convexity/concavity, specifically, g_0 is concave if:

$$\lambda g_0(\mathrm{x}_\mathbf{1}) + (\mathbf{1} - \lambda)\mathbf{g_0}(\mathrm{x}_\mathbf{2}) \leq g_0(\lambda \mathrm{x}_\mathbf{1} + (\mathbf{1} - \lambda)\mathrm{x}_\mathbf{2}), \quad \forall \mathrm{x}_\mathbf{1}, \mathrm{x}_\mathbf{2} \in \mathbb{R}^\mathbf{d} \text{ and } \forall \lambda \in [\mathbf{0}, \mathbf{1}]$$

Furthermore, saying g_0 is monotonically increasing means that

$$\text{if } \mathrm{x}_\mathbf{1} \leq \mathrm{x}_\mathbf{2}, \text{ then } \mathbf{g_0}(\mathrm{x}_\mathbf{1}) \leq \mathbf{g_0}(\mathrm{x}_\mathbf{2}),$$

where the inequality of $\mathrm{x}_\mathbf{1} \leq \mathrm{x}_\mathbf{2}$ means that every component of $\mathrm{x}_\mathbf{2}$ is greater than or equal to the corresponding component of $\mathrm{x}_\mathbf{1}$. Here, we denote G_2 as the set of functions satisfying these constraints.

[1] A variable bandwidth method allows the bandwidth associated with a particular regressor to vary with the density of the data.

2.2 Shape-Constrained Kernel-Weighted Least-Squares (SCKLS) with Local Linear

Given observations $\{\boldsymbol{X}_j, y_j\}_{j=1}^n$, we state the (multivariate) local linear kernel estimator developed by Stone (1977) and Cleveland (1979) as

$$\min_{a,\boldsymbol{b}} \sum_{j=1}^{n} (y_j - a - (\boldsymbol{X}_j - \boldsymbol{x})'\boldsymbol{b})^2 K\left(\frac{\boldsymbol{X}_j - \boldsymbol{x}}{\boldsymbol{h}}\right), \tag{1}$$

where a is a functional estimate, and $\boldsymbol{b}$ is an estimate of the slope of the function at $\boldsymbol{x}$ with $\boldsymbol{x}$ being an arbitrary point in the input space, $K(\frac{\boldsymbol{X}_j - \boldsymbol{x}}{\boldsymbol{h}})$ denotes a product kernel, and $\boldsymbol{h}$ is a vector of bandwidths (see Racine and Li (2004) for more detail). We note that the objective function uses kernel weights, so more weight is given to the observations that are closer to the point $\boldsymbol{x}$.

We introduce a set of m points, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$, for evaluating constraints, which we call evaluation points, and impose shape constraints on the local linear kernel estimator. In the spirit of local linear kernel estimator, we define shape constrained kernel-weighted least squares (SCKLS) estimator, for the case of monotonicity and concavity, to be the function $\hat{g}_n : \mathbb{R}^d \to \mathbb{R}$ such that

$$\hat{g}_n(\boldsymbol{x}; \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}) = \min_{i \in \{1,\ldots,m\}} \left\{\hat{a}_i + (\boldsymbol{x} - \boldsymbol{x}_i)'\hat{\boldsymbol{b}}_i\right\} \tag{2}$$

for any $\boldsymbol{x} \in \mathbb{R}^d$, where $\hat{\boldsymbol{a}} = (\hat{a}_1, \ldots, \hat{a}_m)'$ and $\hat{\boldsymbol{b}} = (\hat{\boldsymbol{b}}_1', \ldots, \hat{\boldsymbol{b}}_m')'$ are the solutions to the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{a},\boldsymbol{b}} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} (y_j - a_i - (\boldsymbol{X}_j - \boldsymbol{x}_i)'\boldsymbol{b}_i)^2 K\left(\frac{\boldsymbol{X}_j - \boldsymbol{x}_i}{\boldsymbol{h}}\right) \\ \text{subject to} \quad & a_i - a_l \geq \boldsymbol{b}_i'(\boldsymbol{x}_i - \boldsymbol{x}_l), \qquad i, l = 1, \ldots, m \\ & \boldsymbol{b}_i \geq 0, \qquad i = 1, \ldots, m. \end{aligned} \tag{3}$$

The first set of constraints in (3) imposes concavity and the second set of constraints imposes nonnegativity of $\boldsymbol{b}_i$ at each evaluation point $\boldsymbol{x}_i$. For more details see Kuosmanen (2008). Note that (2) implies the functional estimate is constructed by taking the minimum of linear interpolations between the evaluation points. This makes SCKLS a globally shape constrained function although it is a non-smooth piecewise linear function.

The SCKLS estimator requires the user to specify the number and the locations of the evaluation points. A standard method for determining the location of evaluation points, $\{\boldsymbol{x}_i\}_{i=1}^m$, is to construct a uniform grid, where each dimension is divided using equal spacing. However, we can address the skewness of input variable distributions common in manufacturing survey data by using a nonuniform grid method, specifically percentile gridding, to specify evaluation points.

Alternatively, we can deal with the input skewness by applying the k-nearest neighbor (k-NN) approach (Li, Liu, and Li, 2017). The k-NN approach uses a smaller bandwidth in dense data regions and a larger bandwidth when the data are sparse. The analysis in Section 6 uses both a percentile grid and k-NN approach to define the kernel function. For details of these extensions, see Appendix A.

As the density of the evaluation points increases, the estimated function potentially has more hyperplane components and is more flexible; however, the computation time typically increases. If a smooth functional estimate is preferred, see Nesterov (2005) and Mazumder et al. (2017), where methods for smoothing are provided. In practice, we propose to select the bandwidth vector $\boldsymbol{h}$ via the leave-one-out cross-validation based on the unconstrained estimator. See Section 5 for the details.

Appendix A proposes several alternative implementations of the SCKLS estimator: (1) SCKLS with local polynomial approximation, (2) a k-nearest neighbor (k-NN) approach, and (3) nonuniform grid method.

3. THEORETICAL PROPERTIES OF SCKLS

For mathematical concreteness, we next consider the statistical properties of SCKLS under monotonicity and concavity constraints. Recall that G_2 is the class of functions which are monotonically increasing and globally concave, and g_0 is the truth to be estimated from n pairs of observations. We make the following assumptions:

Assumption 1.

(i) $\{\boldsymbol{X}_j, y_j\}_{j=1}^{\infty}$ are a sequence of iid random variables with $y_j = g_0(\boldsymbol{X}_j) + \epsilon_j$.

(ii) $g_0 \in G_2$ and is twice-differentiable.

(iii) $\boldsymbol{X}_j$ follows a distribution with continuous density function f and support $\boldsymbol{S}$. Here, $\boldsymbol{S}$ is a convex, nondegenerate and compact subset of $\mathbb{R}^d$. Moreover,

$$\min_{\boldsymbol{x} \in \boldsymbol{S}} f(\boldsymbol{x}) > 0.$$

(iv) The conditional probability density function of ϵ_j, given $\boldsymbol{X}_j$, denoted as $p(e|\boldsymbol{x})$, is continuous with respect to both e and $\boldsymbol{x}$, with the mean function

$$\mu(\cdot) = E(\epsilon_j | \boldsymbol{X}_j = \cdot) = 0$$

and the variance function

$$\sigma^2(\cdot) = \mathrm{Var}(\epsilon_j | \boldsymbol{X}_j = \cdot)$$

bounded away from 0 and continuous over $\boldsymbol{S}$. Moreover, $\sup_{\boldsymbol{x} \in \boldsymbol{S}} E(\epsilon_j^4 | \boldsymbol{X}_j = \boldsymbol{x}) < \infty$.

(v) $K(\cdot)$ is a nonnegative, Lipschitz second-order kernel with a compact and convex support. For simplicity, we set the bandwidth associated with each explanatory variable, h_k, for $k = 1, \ldots, d$, to be $h_1 = \cdots = h_d = h$.

(vi) $h = O(n^{-1/(4+d)})$ as $n \to \infty$.

Here, (i) states that the data are iid; (ii) says that the constraints we impose on the SCKLS estimator are satisfied by the true function; (iii) makes a further assumption on the distribution of the covariates; (iv) states that the noise can be heteroscedastic in certain ways, but requires the change in the variance to be smooth; (v) is rather standard in local polynomial estimation to facilitate the theoretical analysis; and (vi) assures the bandwidths become sufficiently small as $n \to \infty$ so that both the bias and the variance from local averaging go to zero. For details of the consistency of local linear estimator and a discussion of some of these conditions, see Masry (1996), Li and Racine (2007), and Fan and Guerre (2016).

We consider two scenarios: let the number of evaluation points (denoted by m) grow with n, or fix the number of evaluation points a priori. For simplicity, we also assume that the evaluation points are drawn independent of $\{\boldsymbol{X}_j, y_j\}_{j=1}^n$.

Assumption 2.

(i) The number of evaluation points $m \to \infty$ as $n \to \infty$. For simplicity, we assume that the empirical distribution of $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ converges to a distribution Q that has support $\boldsymbol{S}$ (i.e., as defined in Assumption 1(iv))) and a continuous differentiable density function $q : \boldsymbol{S} \to \mathbb{R}$ satisfying $\min_{\boldsymbol{x}\in\boldsymbol{S}} q(\boldsymbol{x}) > 0$.

(ii) The number of evaluation points m is fixed. All the evaluation points lie in the interior of $\boldsymbol{S}$. Moreover,

$$\frac{\sup_{\boldsymbol{x}\in\boldsymbol{S}} \min_{i=1,\ldots,m} \|\boldsymbol{x} - \boldsymbol{x}_i\|}{\min_{i\neq j; i,j\in\{1,\ldots,m\}} \|\boldsymbol{x}_j - \boldsymbol{x}_i\|} \leq \kappa$$

for some $\kappa \geq 1$ (i.e., $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ are reasonably well spread across $\boldsymbol{S}$).

Our main results are summarized below. A short discussion on our proof strategy and the proofs are available in Appendix B.

Theorem 1. Suppose that Assumption 1(i)–1(vi) and Assumption 2(i) or 2(ii) hold. Then,

$$\frac{1}{m}\sum_{i=1}^m \{\hat{g}_n(\boldsymbol{x}_i) - g_0(\boldsymbol{x}_i)\}^2 = O_p(n^{-4/(4+d)} \log n).$$

Theorem 2.

1. (The case of an increasing m) Suppose that Assumption 1(i)–1(vi) and Assumption 2(i) hold. Let $\boldsymbol{C}$ be any fixed closed set that belongs to the interior of $\boldsymbol{S}$. Then with probability one, as $n \to \infty$, the SCKLS estimator satisfies
$$\sup_{\boldsymbol{x}\in\boldsymbol{C}} \left|\hat{g}_n(\boldsymbol{x}) - g_0(\boldsymbol{x})\right| \to 0.$$
2. (The case of a fixed m) Suppose that Assumption 1(i)–1(vi) and Assumption 2(ii) hold. Then, as $n \to \infty$, with probability one, the estimates from SCKLS satisfy
$$\hat{a}_i \to g_0(\boldsymbol{x}_i) \quad \text{and} \quad \hat{\boldsymbol{b}}_i \to \frac{\partial g_0}{\partial \boldsymbol{x}}(\boldsymbol{x}_i)$$
for all $i = 1, \ldots, m$.

Note that this convergence rate is nearly optimal (differing only by a factor of $\log n$). However, in the above, we only manage to show that the SCKLS estimator converges at the evaluation points or in the interior of the domain. It is known that shape-constrained estimators tend to suffer from bad boundary behaviors. For instance, the quantity $\sup_{\boldsymbol{S}} |\hat{g}_n^{\text{CNLS}}(\boldsymbol{x}) - g_0(\boldsymbol{x})|$ does *not* converge to zero in probability, where $\hat{g}_n^{\text{CNLS}}$ is the CNLS estimator. Though for SCKLS, if we let the number of evaluation points, m, grow at a rate slower than n, we argue that we can both alleviate the boundary inconsistency and improve the computational efficiency.

Assumption 3. The number of evaluation points $m = o(n^{2/(4+d)}/\log n)$ as $n \to \infty$.

Theorem 3. Suppose that Assumption 1(i)–1(vi), Assumption 2(i) and Assumption 3 hold. Then, with probability one, as $n \to \infty$, the SCKLS estimator satisfies

$$\sup_{\boldsymbol{x}\in\boldsymbol{S}} |\hat{g}_n(\boldsymbol{x}) - g_0(\boldsymbol{x})| \to 0.$$

We also note that CNLS can be viewed as a special case of SCKLS when we let the set of evaluation points be $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ and the bandwidth vector $\|\boldsymbol{h}\| \to \boldsymbol{0}$. See Appendix A for the proof of the relationship between CNLS and SCKLS, together with a further discussion on the relationship between SCKLS and alternative shape constrained estimators such as CWB.

4. SHAPE MISSPECIFICATION: THEORY AND TESTING

4.1 Misspecification of the Shape Restrictions

So far we have assumed in our estimation procedures that $g_0 \in G_2$, where G_2 is the class of functions which are monotonically increasing and globally concave. To understand the behavior of SCKLS, we are interested in its performance when $g_0 \notin G_2$.

Let Q be a distribution on $\boldsymbol{S}$ (as in Assumption 2(i)) and define $g^* : \boldsymbol{S} \to \mathbb{R}$ as

$$g_0^* := \operatorname*{argmin}_{g\in G_2} \int_{\boldsymbol{S}} \{g(\boldsymbol{x}) - g_0(\boldsymbol{x})\}^2 Q(d\boldsymbol{x}).$$

The existence and Q-uniqueness of g_0^* follows from the well-known results about the projection onto a cone in the Hilbert space. When $g_0 \in G_2$, it is easy to check that $g_0^* = g_0$. See also Lim and Glynn (2012). The following result can be viewed as a generalization of Theorem 2.

Theorem 4. Suppose that Assumption 1(i), 1(iii)–1(vi) and Assumption 2(i) hold. Furthermore, suppose that g_0 is twice-differentiable. Let $\boldsymbol{C}$ be any compact set that belongs to the interior of $\boldsymbol{S}$. Then with probability one, as $n \to \infty$, the SCKLS estimator satisfies

$$\sup_{\boldsymbol{x}\in\boldsymbol{C}} \left|\hat{g}_n(\boldsymbol{x}) - g_0^*(\boldsymbol{x})\right| \to 0.$$

Theorem 4 assures us that the SCKLS estimator converges uniformly on a compact set to the function g_0^* that is closest in L^2 distance to the true function g_0 for which our estimator is misspecified. Consequently, as long as g_0 is not too far away from G_2, our estimator can still be used as a reasonable approximation to the truth, especially when the sample size is moderate. See Appendix E for a numerical demonstration.

4.2 Hypothesis Testing for the Shape

Admittedly, the SCKLS estimator can be inappropriate if the shape constraints are not fulfilled by g_0. Thus, we propose a procedure based on the SCKLS estimators for testing

$$H_0 : \{g_0 : \boldsymbol{S} \to \mathbb{R}\} \in G_2 \quad \text{against} \quad H_1 : \{g_0 : \boldsymbol{S} \to \mathbb{R}\} \notin G_2.$$

Denote by

$$\tilde{r}^2\left(\{\boldsymbol{X}_j, y_j\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\right) = \min_{\boldsymbol{a},\boldsymbol{b}} \sum_{i=1}^m \sum_{j=1}^n (y_j - a_i - (\boldsymbol{X}_j - \boldsymbol{x}_i)'\boldsymbol{b}_i)^2 K\left(\frac{\boldsymbol{X}_j - \boldsymbol{x}_i}{\boldsymbol{h}}\right);$$

the value of the objective function that is minimized by the local linear kernel estimator. And denote by

$$\hat{r}^2\big(\{\boldsymbol{X}_j, y_j\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\big) = \min_{\boldsymbol{a},\boldsymbol{b}} \sum_{i=1}^m \sum_{j=1}^n (y_j - a_i - (\boldsymbol{X}_j - \boldsymbol{x}_i)'\boldsymbol{b}_i)^2 K\left(\frac{\boldsymbol{X}_j - \boldsymbol{x}_i}{\boldsymbol{h}}\right),$$

$$\text{subject to } a_i - a_l \geq \boldsymbol{b}_i'(\boldsymbol{x}_i - \boldsymbol{x}_l) \text{ and } \boldsymbol{b}_i \geq 0, \quad i, l = 1, \ldots, m.$$

Here, $\hat{r}^2(\cdot,\cdot)$ is the value of the objective function that is minimized by SCKLS.

We focus on the test statistic

$$T_n := T\big(\{\boldsymbol{X}_j, y_j\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\big) = \left[\frac{1}{mnh^d}\big\{\hat{r}^2\big(\{\boldsymbol{X}_j, y_j\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\big) - \tilde{r}^2\big(\{\boldsymbol{X}_j, y_j\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\big)\big\}\right]^{1/2},$$

which is a rescaled version of the difference between the values of the same objective function (with the same bandwidth $\boldsymbol{h}$), optimized either with or without the shape constraints. Intuitively, the value of this statistic should be small if $g_0 \in G_2$. This statistic can also be viewed as a smoothed and rescaled version of the goodness-of-fit statistic.

Here, we focus on the boundary case when g_0 is constant (i.e., $g_0 = 0$) because it is hardest to evaluate the null hypothesis when g_0 is both nonincreasing and nondecreasing and both concave and convex, intuitively and theoretically and it allows us to control the size of our test statistic. Since the noise here might be non-homogeneous, we use the wild bootstrap to approximate the distribution of the test statistic under H_0. See Wu (1986), Liu (1988), Mammen (1993), and Davidson and Flachaire (2008) for an overview of the wild bootstrap procedure.

Our testing procedure has three steps:

(1) Estimate the error at each $\boldsymbol{X}_j$ by $\tilde{\epsilon_j} = y_j - \tilde{g}_n(\boldsymbol{X}_j)$ for $j = 1, \ldots, n$, where $\tilde{g}$ is the unconstrained local linear estimator with kernel and bandwidth satisfying Assumptions 1(v)–(vi).

(2) The wild bootstrap method is used to construct a critical region for T_n. Let B be the number of Monte Carlo iterations. For every $k = 1, \ldots, B$, let $\boldsymbol{u}_k = (u_{1k}, \ldots, u_{nk})'$ be a random vector with components sampled independently from the Rademacher distribution, that is, $P(u_{jk} = 1) = P(u_{jk} = -1) = 0.5$. Furthermore, let $y_{jk} = u_{jk}\ \tilde{\epsilon}_j$. Then, the wild bootstrap test statistic is

$$T_{nk} = T\big(\{\boldsymbol{X}_j, y_{jk}\}_{j=1}^n, \{\boldsymbol{x}_i\}_{i=1}^m\big).$$

(3) Define the Monte Carlo p-value as[2]

$$p_n = \frac{1}{B}\sum_{k=1}^B \mathbf{1}_{\{T_n \leq T_{nk}\}}.$$

For a test of size $\alpha \in (0, 1)$, we reject H_0 if $p_n < \alpha$.

[2]Since we underestimate the level of the errors in Step 1 by a factor of roughly $n^{-2/(4+d)}$, for the theoretical development, we address this bias issue by modifying the p-value to be $p_n = \frac{1}{B}\sum_{k=1}^B \mathbf{1}_{\{T_n \leq T_{nk}+\Delta_n\}}$, where $\Delta_n = O(n^{-2/(4+d)}\log n)$. Note that if we fix m and pick $h = O(n^{-\eta})$ for $\eta \in (\frac{1}{4+d}, \frac{1}{d})$, then $\Delta_n / T_{nk} = o_p(1)$ as $n \to \infty$, that is, this correction has a negligible effect. Indeed, our experience suggests that this modification offers little improvement in terms of finite sample performance in our simulation study.

A few remarks are in order.

First, here we conveniently implemented the simplest wild bootstrap scheme to simplify our analysis, in line with the work of Davidson and Flachaire (2008). Instead of imposing the Rademacher distribution on u_{kj}, we can also use any distribution with zero-mean and unit-variance. One popular choice suggested by Mammen (1993) is

$$u_{jk} = \begin{cases} -\frac{\sqrt{5}-1}{2} & \text{with probability } \frac{5+\sqrt{5}}{10} \\ \frac{\sqrt{5}+1}{2} & \text{with probability } \frac{5-\sqrt{5}}{10} \end{cases}.$$

Second, note that the definition of y_{jk} in Step 2 makes this a test of the residuals, that is, when drawing bootstrap samples, we use $y_{jk} = u_{jk}\ \tilde{\epsilon}_j$ instead of $y_{jk} = \hat{g}_n(\boldsymbol{X}_j) + u_{jk}\ \tilde{\epsilon}_j$. From this perspective, our test is similar to the univariate monotonicity test in Hall and Heckman (2000). One reason behind this choice is to avoid the boundary inconsistency of the bootstrap procedure. See Andrews (2000) and Cavaliere, Nielsen, and Rahbek (2017) who addressed this issue in a much simpler setup. Generally speaking, testing the null hypothesis becomes harder when g_0 is on the boundary of G_2. In practice, we could use $y_{jk} = \hat{g}_n(\boldsymbol{X}_j) + u_{jk}\ \tilde{\epsilon}_j$ in certain scenarios (e.g., when testing g_0 is a strictly increasing and strictly concave function against $g_0 \notin G_2$), and slight improvements are observed in terms of finite-sample performance.

We now look into the theoretical properties of our procedure under both H_0 and H_1. See Appendix B for the proof.

Theorem 5. Suppose that Assumptions 1(i), (iii)–(v) and 2(i) hold, and the conditional error distribution (i.e., $\epsilon_j|\boldsymbol{X}_j$) is symmetric. Furthermore, assume that g_0 is continuously twice-differentiable and let $h = O(n^{-\eta})$ for some fixed $\eta \in (\frac{1}{4+d}, \frac{1}{d})$. Let $B := B(n) \to \infty$ as $n \to \infty$. Then, for any given $\alpha \in (0, 1)$,

- Type I error: for any $g_0 \in G_2$, $\limsup_{n\to\infty} P(p_n < \alpha) \leq \alpha$;
- Type II error: for any $g_0 \notin G_2$, $\limsup_{n\to\infty}\{1 - P(p_n < \alpha)\} = 0$.

In addition, if we replace Assumption 2(i) by Assumption 2(ii), the same conclusions hold for sufficiently large m.

See also Section 5 for the finite-sample performance of our test in a simulation study, where we demonstrate that the proposed test controls both Type I and Type II errors reasonably well. Additionally, Appendix C describes our procedure for testing affinity using SCKLS.

5. SIMULATION STUDY

5.1 Numerical Experiments on Estimation

5.1.1 The Setup. We now examine the finite-sample performance and robustness of the proposed estimator through Monte Carlo simulations. We run our experiments on a computer with Intel Core2 Quad CPU 3.00 GHz and 8 GB RAM. We compare the performance of SCKLS is compared with that of CNLS and LL. See Appendix E for a comparisons of SCKLS with CWB. For the SCKLS and the CNLS estimator, we solve the quadratic programming problems with MATLAB using the built-in quadratic programming solver, `quadprog`. We run two sets of experiments varying the number of observations (n), the

number of evaluation points (m), and the number of the inputs (d).

We also run additional experiments to show the robust performance of the SCKLS estimator under alternative conditions. See Appendix E for the results.

We measure the estimator's performance using root mean squared errors (RMSE) based on two criteria: the distance from the estimated function to the true function measured (1) at the observed points and (2) at the evaluation points constructed on an uniform grid, respectively. As CNLS estimates hyperplanes at observation points, we use linear interpolation to obtain the RMSE of CNLS[3]. We replicate each scenario 10 times and report the average and standard deviation.

5.1.2 Tuning Parameter Selection. For the SCKLS estimator, we use the Gaussian kernel function $K(\cdot)$ and leave-one-out cross-validation (LOOCV) for bandwidth selection. LOOCV is a data-driven method, and has been shown to perform well for unconstrained kernel estimators such as local linear (Stone 1977). We apply LOOCV procedure on unconstrained estimates (i.e., local linear) to select the bandwidth for SCKLS to reduce the computational burden and because SCKLS is relatively insensitive to the bandwidth choice (see for example Experiment 1. Fixed Number of Evaluation Points). For further computational improvements, we apply the iterative algorithm described in Appendix D .

5.1.3 Results.

5.1.3.1 Experiment 1. Fixed Number of Evaluation Points. We consider a Cobb–Douglas production function with d-inputs and one-output, $g_0(x_1, \ldots, x_d) = \prod_{k=1}^{d} x_k^{\frac{0.8}{d}}$. For each pair $(\boldsymbol{X}_j, y_j)$, each component of the input, $\boldsymbol{X}_{jk}$, is randomly and independently drawn from uniform distribution $unif[1, 10]$, and the additive noise, ϵ_j, is randomly sampled from a normal distribution, $N(0, 0.7^2)$. We consider 15 different scenarios with different numbers of observations (100, 200, 300, 400 and 500) and input dimensions (2, 3, and 4). The structure and data generation process of Experiment 1 follows Lee et al. (2013). We fix the number of evaluation points at approximately 400 and locate them on a uniform grid.

For this experiment, we compare the following four estimators: SCKLS, CNLS, local linear kernel (LL), and parametric Cobb–Douglas estimator. The latter estimator serves as a baseline because it is correctly specified parametric form. Tables 1 and 2 show for Experiment 1 the RMSE measured on observation points and evaluation points, respectively. The number in parentheses is the standard deviation of RMSE values computed by 10 replications. Note the standard derivations are generally small compared to the parameter estimates, which indicates low variability even after only 10 replications. A more extensive set of results for this experiment is summarized in Appendix E. The SCKLS estimator has the lowest RMSE in most scenarios even when RMSE is measured on observation points (note that the SCKLS estimator imposes the global shape constraints via evaluation points in Equation (3)). Also as expected, the performance of SCKLS estimator improves as the number of observation points increases. Moreover, the SCKLS estimator

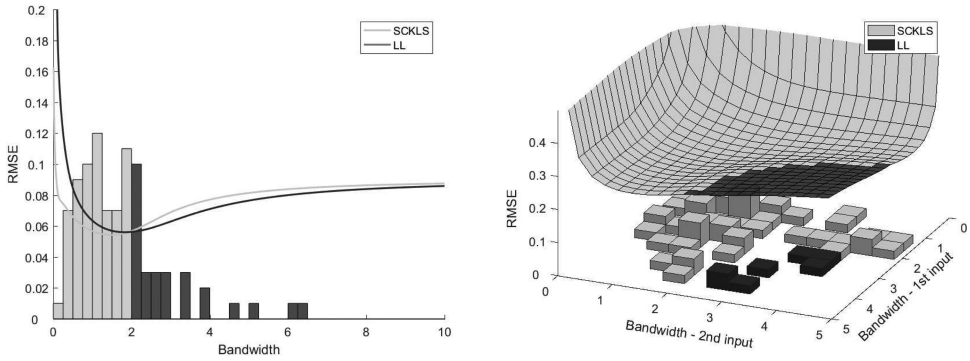
[3]The CNLS estimates include the second stage linear programming estimation procedure described by Kuosmanen and Kortelainen (2012) to find the minimum extrapolated production function.

Table 1. RMSE on observation points for Experiment 1

		Average of RMSE on observation points				
Number of observations		100	200	300	400	500
2-input	SCKLS	**0.193**	0.171	0.141	**0.132**	0.118
		(0.053)	(0.047)	(0.032)	(0.029)	(0.017)
	CNLS	0.229	**0.163**	**0.137**	0.138	**0.116**
		(0.042)	(0.037)	(0.010)	(0.027)	(0.016)
	LL	0.212	0.166	0.149	0.152	0.140
		(0.079)	(0.042)	(0.028)	(0.028)	(0.028)
	Cobb–Douglas	0.078	0.075	0.048	0.039	0.043
3-input	SCKLS	**0.230**	**0.187**	**0.183**	**0.152**	**0.165**
		(0.050)	(0.026)	(0.032)	(0.019)	(0.031)
	CNLS	0.294	0.202	0.189	0.173	0.168
		(0.048)	(0.035)	(0.020)	(0.014)	(0.020)
	LL	0.250	0.230	0.235	0.203	0.181
		(0.068)	(0.050)	(0.052)	(0.050)	(0.021)
	Cobb–Douglas	0.104	0.089	0.070	0.047	0.041
4-input	SCKLS	**0.225**	**0.248**	**0.228**	**0.203**	**0.198**
		(0.038)	(0.020)	(0.037)	(0.042)	(0.028)
	CNLS	0.315	0.294	0.246	0.235	0.214
		(0.039)	(0.027)	(0.024)	(0.029)	(0.015)
	LL	0.256	0.297	0.252	0.240	0.226
		(0.044)	(0.057)	(0.056)	(0.060)	(0.038)
	Cobb–Douglas	0.120	0.073	0.091	0.067	0.063

Table 2. RMSE on evaluation points for Experiment 1

		Average of RMSE on evaluation points				
Number of observations		100	200	300	400	500
2-input	SCKLS	**0.219**	0.189	**0.150**	**0.147**	**0.128**
		(0.053)	(0.057)	(0.034)	(0.030)	(0.021)
	CNLS	0.350	0.299	0.260	0.284	0.265
		(0.082)	(0.093)	(0.109)	(0.119)	(0.078)
	LL	0.247	**0.182**	0.167	0.171	0.156
		(0.101)	(0.053)	(0.030)	(0.030)	(0.034)
	Cobb–Douglas	0.076	0.076	0.049	0.040	0.043
3-input	SCKLS	**0.283**	**0.231**	**0.238**	**0.213**	**0.215**
		(0.072)	(0.033)	(0.030)	(0.029)	(0.034)
	CNLS	0.529	0.587	0.540	0.589	0.598
		(0.112)	(0.243)	(0.161)	(0.109)	(0.143)
	LL	0.336	0.340	0.360	0.326	0.264
		(0.085)	(0.093)	(0.108)	(0.086)	(0.042)
	Cobb–Douglas	0.116	0.098	0.080	0.052	0.046
4-input	SCKLS	**0.321**	**0.357**	**0.329**	**0.308**	**0.290**
		(0.046)	(0.065)	(0.049)	(0.084)	(0.044)
	CNLS	0.845	0.873	0.901	0.827	0.792
		(0.188)	(0.137)	(0.151)	(0.235)	(0.091)
	LL	0.482	0.527	0.483	0.495	0.445
		(0.115)	(0.125)	(0.146)	(0.153)	(0.074)
	Cobb–Douglas	0.146	0.091	0.115	0.081	0.080



(a) One-input (b) Two-input

Figure 1. The histogram shows the distribution of bandwidths selected by LOOCV. The curves show the relative performance of each estimator.

performs better than the LL estimator particularly in higher dimensional functional estimation. This provides empirical evidence that the shape constraints in SCKLS are helpful in improving the finite-sample performance as compared to LL. Note that LL appears to have larger RMSE values on evaluation points which are located in input space regions with sparse observations. This implies that the SCKLS estimator has more robust out-of-sample performance than the LL estimator due to the shape constraints. We also observe that the performance of the CNLS estimator measured at the evaluation points is worse than that measured at the observations. CNLS often has ill-defined hyperplanes which are very steep/shallow at the edge of the observed data, and this over-fitting leads to poor out-of-sample performance. In contrast, the SCKLS estimator performs similarly for both the observation points and evaluation points, because the construction of the grid that completely covers the observed data makes the SCKLS estimator more robust.

We also conduct simulations with different bandwidths to analyze the sensitivity of each estimator to bandwidths. We compare SCKLS and LL with bandwidth $h \in [0, 10]$ with an increment by 0.01 for the 1-input setting, and we use bandwidth $\boldsymbol{h} \in [0, 5] \times [0, 5]$ with an increment by 0.25 in each coordinate for the 2-input setting. We simulate 100 datasets to compute the RMSE for each bandwidth as well as for the bandwidth via LOOCV. Figure 1 displays the average RMSE of each estimator. The histogram shows the distribution of bandwidths selected by LOOCV. The instances when SCKLS and LL provide the lowest RMSE are shown in light gray and dark gray, respectively. For the one-input scenario, the SCKLS estimator performs better than the LL estimator for bandwidth between 0.25 and 2.25 as shown in (a). For the two-input scenario, the SCKLS estimator performs better for most of the LOOCV values as shown by the majority of the histogram colored in light gray. This indicates that LOOCV, calculated using the unconstrained estimator, provides bandwidths that work well for the SCKLS estimator. Importantly, the SCKLS estimator does not appear to be very sensitive to the bandwidth selection method since, heuristically, the shape constraints help reduce the variance of the estimator. Finally, we note that similar results can be obtained in experimental settings with lower signal-to-noise level, or with non-uniform input. See Appendix E for more details.

5.1.3.2 Experiment 2. Different Numbers of Evaluation Points. The setting is the same as Experiment 1. However, now we consider 9 different scenarios with different numbers of evaluation points (100, 300, and 500) and input dimensions (2, 3, and 4). We fix the number of observed points at 400.

We show the performance of SCKLS. Tables 3 and 4 shows for Experiment 2 the RMSE measured on observations and evaluation points respectively. Both tables show that empirically even if we increase the number of evaluation points, the RMSE value does not change significantly. This has important implications for the running time. Specifically, we can reduce the calculation time by using a rough grid without sacrificing too much in terms of RMSE performance of the estimator.

Table 3. RMSE on observation points for Experiment 2

		Average of RMSE on observation points		
Number of evaluation points		100	300	500
2-input	SCKLS	0.142	0.141	0.141
3-input	SCKLS	0.198	0.203	0.197
4-input	SCKLS	0.239	0.207	0.206

Table 4. RMSE on evaluation points for Experiment 2.

		Average of RMSE on evaluation points		
Number of evaluation points		100	300	500
2-input	SCKLS	0.181	0.164	0.158
3-input	SCKLS	0.304	0.267	0.257
4-input	SCKLS	0.383	0.296	0.270

5.2 Numerical Experiments on Testing the Imposed Shape

Experiment 3. We test monotonicity and concavity for data generated from the following single-input and single-output DGP:

$$g_0(x) = x^p \tag{4}$$

and

$$g_0(x) = \frac{1}{1 + \exp(-5\log(2x))}. \tag{5}$$

With n observations, for each pair (X_j, y_j), each input, X_j, is randomly and independently drawn from uniform distribution unif [0, 1]. In this simulation, we use the following multiplicative noise to validate whether the wild bootstrap can handle nonhomogeneous noise.

$$y_j = g_0(X_j) + (X_j + 1) \cdot \epsilon_j,$$

where ϵ_j, is randomly and independently sampled from a normal distribution, $N(0, \sigma^2)$. We use three different DGP scenarios A, B, and C. For scenarios A and B, we use function (4) where the exponent parameter p defines whether the function g_0 is an element of the class of functions G_2 or not. We use $p = \{0, 2\}$ for scenarios A and B, respectively, where $g_0 \in G_2$ if $p = 0$, and $g_0 \notin G_2$ if $p = 2$ since g_0 is strictly convex. For scenario C, we consider an "S"-shape function defined by (5) which violates both global concavity and convexity. We consider different sample sizes $n = \{100, 300, 500\}$ and standard deviation of the noise $\sigma = \{0.1, 0.2\}$, and perform 500 simulations to compute the rejection rate for each scenario. We assume that we do not know the distribution of the noise in advance and use the wild bootstrap procedure described in Section 4.2 with $B = 200$.

Table 5 shows the rejection rate for each DGP. For high signal-to-noise ratio scenarios ($\sigma = 0.1$), the test works well even with a small-sample size. Our test is able to control the Type I error, as illustrated in scenario A. In addition, the Type II error of our test is small for the scenarios B and C where shape constraints are violated by the DGP. Furthermore, for low signal-to-noise ratio scenarios ($\sigma = 0.2$), the rejection rate for scenarios B and C significantly improves when the sample size is increased from 100 to 300. Indeed, for larger noise scenarios more data is required for the test to have power. Thus, our test seems informative enough to guide users to avoid imposing shape constraints on the data generated from misspecified functions.

Table 5. Rejection rate (%) of the test for monotonicity and concavity

		Power of the Test (α)			
Sample size (n)	DGP Scenario	0.05	0.01	0.05	0.01
		$\sigma = 0.1$		$\sigma = 0.2$	
100	A (H_0)	5.8	2.0	8.0	2.6
	B (H_1)	98.6	94.6	55.0	36.2
	C (H_1)	98.6	94.4	42.6	24.2
300	A (H_0)	6.8	1.8	6.6	3.0
	B (H_1)	100.0	100.0	92.0	83.2
	C (H_1)	100.0	100.0	97.0	86.8
500	A (H_0)	5.4	1.6	5.6	1.4
	B (H_1)	100.0	100.0	99.4	97.2
	C (H_1)	100.0	100.0	99.8	99.4

6. APPLICATION

We apply the proposed method to estimate the production function for two large industries in Chile: plastic (2520) and wood manufacturing (2010), where the values inside the parentheses indicate the CIIU3 industry code. There are some existing studies which analyze the productivity of Chilean data; see, for example, Pavcnik (2002), who analyzed the effect of trade liberalization on productivity improvements. Other researchers have analyzed the productivity of Chilean manufacturing including Benavente (2006), Alvarez and Görg (2009), and Levinsohn and Petrin (2003). However, the above-cited work use strong parametric assumptions and older data. Most studies use the Cobb–Douglas functional form which restricts the elasticity of substitution to be 1. When diminishing marginal productivity of inputs characterizes the data, the Cobb–Douglas functional form imposes that the most productive scale size is at the origin. We relax the parametric assumptions and estimate a shape constrained production function nonparametrically using data from 2010. We examine the marginal productivity, marginal rate of substitution, and most productive scale size (MPSS) to analyze the structure of the industries. We also investigate how productivity differs between exporting and nonexporting firms, as exporting has become an important source of revenue in Chile[4]. See Appendix G for the details of estimation and comparison across different estimators.

6.1 The Census of Chilean Manufacturing Plants

We use the Chilean Annual Industrial Survey provided by Chile's National Institute of Statistics[5]. The survey covers manufacturing establishments with ten or more employees. We define Capital and Labor as the input variables and Value Added as the output variable of the production function[6]. Capital and Value Added are measured in millions of Chilean peso while Labor is measured as the total man-hours per year. We use cross-sectional data from the plastic and the wood industries.

Many researchers have found positive effects of exporting for other countries using parametric models. See for instance, De Loecker (2007) and Bernard and Jensen (2004). Here, we use SCKLS to relax the parametric assumption for the production

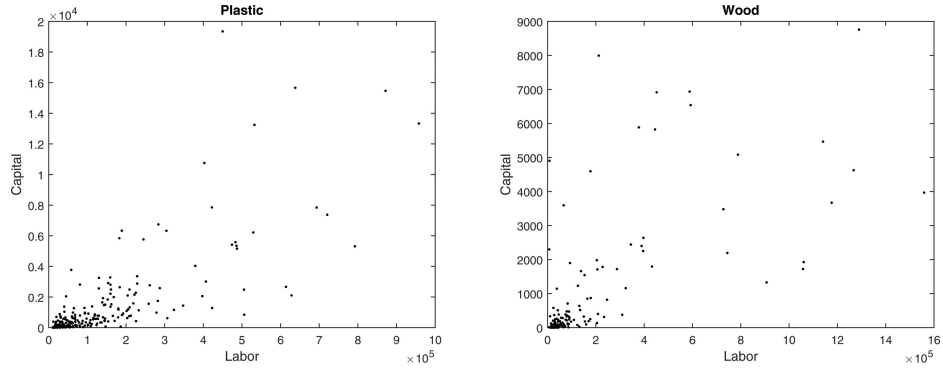
[4]Note that firms' decisions, that is, selecting labor and capital levels with considerations for productivity levels or whether to export, are potentially endogenous. Solutions to this issue are to instrument or build a structural model based on timing assumptions. Our estimator can be embedded within the estimation procedures such as those described in Ackerberg, Caves, and Frazer (2015) to address this issue.

[5]The data are available at *http://www.ine.cl/estadisticas/economicas/manufactura* .

[6]The definition of Labor includes full-time, part-time, and outsourced labors. Capital is defined as a sum of the fixed assets balance such as buildings, machines, vehicles, furniture, and technical software. Value added is computed by subtracting the cost of raw materials and intermediate consumption from the total amount produced. Further details are available at *http://www.ine.cl/estadisticas/economicas/manufactura* .

Table 6. Statistics of Chilean manufacturing data

	Non-exporters ($n = 173$)			Exporters ($n = 72$)			
Plastic (2520)	Labor	Capital (million)	Value Added (million)	Labor	Capital (million)	Value Added (million)	Share of Exports
mean	92155	725.85	546.93	240890	2859	1733.9	0.147
median	55220	258.41	247.05	180330	1329.1	1054.9	0.0524
std	106530	1574	1068.1	212480	3840.2	1678.8	0.201
skewness	3.301	5.2052	5.9214	1.3681	2.4594	1.0678	− 0.303
	Non-exporters ($n = 97$)			Exporters ($n = 35$)			
Wood (2010)	Labor	Capital	Value Added	Labor	Capital	Value Added	Share of Exports
mean	76561	364.93	334.83	501470	3063.4	4524.1	0.542
median	44087	109.48	115.39	378000	2195.4	2673.5	0.648
std	78057	702.35	555.87	436100	2510.3	4466.3	0.355
skewness	2.243	3.5155	3.432	0.81454	0.63943	1.0556	− 0.303



(a) Plastic (2520) (b) Wood (2010)

Figure 2. Labor and Capital of each industry.

function. To capture the effects of exporting, we use a semi-parametric modeling extension of SCKLS. The partially linear model is represented as follows:

$$y_j = \boldsymbol{Z}_j'\boldsymbol{\gamma} + g_0(\boldsymbol{X}_j) + \epsilon_j, \tag{6}$$

where $\boldsymbol{Z}_j = (Z_{j1}, Z_{j2})'$ denotes contextual variables and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ is the coefficient of contextual variables. We model exporting with two variables: a dummy variable indicating the establishments that are exporting and the share of output being exported. For more details see Appendix F.

Table 6 presents the summary of statistics for each industry by exporter/non-exporter. We find that exporters are typically larger than nonexporter in terms of labor and capital. Input variables are positively skewed, indicating there exist many small and few large establishments. Since SCKLS with variable bandwidth (k-nearest neighbor) and non-uniform grid performed the best in our simulation scenarios with non-uniform input data (as indicated in Appendix E), we use these options. We choose the smoothing parameter k via leave-one-out cross-validation. Appendix A explains the details of our implementation of K-NN for the SCKLS estimator.

Figure 2 shows a plot of labor and capital for each industry and shows input data is sparse for large establishments. Beresteanu (2005) proposed to include shape constraints only for the evaluation points that are close to the observations. Thus, in addition to using a percentile grid of evaluation points, we propose to use the evaluation points that are inside the convex hull of observed input $\{\boldsymbol{X}_j\}_{j=1}^n$. See Appendix G for details.

We begin by testing if the Cobb–Douglas production function is appropriate for our data. We use the hypothesis test for correct parametric specification described in Henderson and Parmeter (2015)[7]. The resulting p-value is 0.092 for the plastic industry and 0.007 for the wood industry, respectively. Therefore, the Cobb–Douglas parametric specification is likely to be wrong, particularly applied to the wood industry.

Next, we apply the test proposed in Section 4.2 to determine if imposing global concavity and monotonicity shape constraints is appropriate. We estimate a p-value of 0.302 for the plastic industry and 0.841 for the wood industry, respectively. For both industries, the estimated p-value is not small enough to reject H_0.

[7]We apply a Cobb–Douglas OLS to the second stage data $\{\boldsymbol{X}_j, y_j - \boldsymbol{Z}_j\boldsymbol{\gamma}\}_{j=1}^n$ which removes the effect of contextual variables from observed output. See Appendix F for details.

Table 7. SCKLS fitting statistics for cross-sectional data

Industry	Number ofobservations	R^2
Plastic	245	71.1%
Wood	132	43.8%

6.2 Estimated Production Function and Interpretation

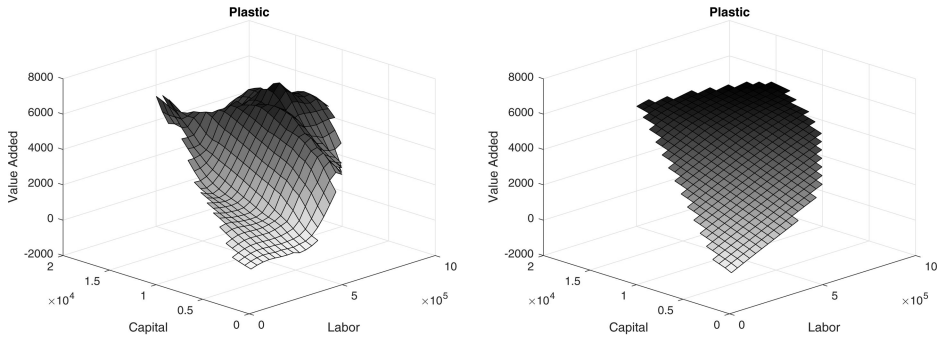
We estimate a semiparametric model with a nonparametric shape constrained production function, a linear model for exporting share of sales, and a dummy variable for exporting. Table 7 shows the goodness of fit (R^2) of the production function: 71.1% of variance is explained in the plastic industry, while 43.8% of variance is explained in the wood industry.

Table 8 reports additional information characterizing the production function: the marginal productivity and the marginal rates of substitution at the 10, 25, 50, 75, and 90 percentiles are reported for both measures. Here, the rate of substitution indicates how much labor is required to maintain the same level of output when we decrease a unit of capital. When comparing the two industries, we find that the wood industry has a larger marginal rate of substitution than the plastic industry. This indicates that capital is more critical in the wood industry than the plastic industry.

We also compare the estimated production function by the local linear and the SCKLS estimators. Figures 3 and 4 show the estimated production function within the convex hull of observations for plastic and wood industries, respectively. Visually, the production function estimated by the LL estimator is difficult to interpret and the values of important economic quantities such as marginal products and marginal rates of

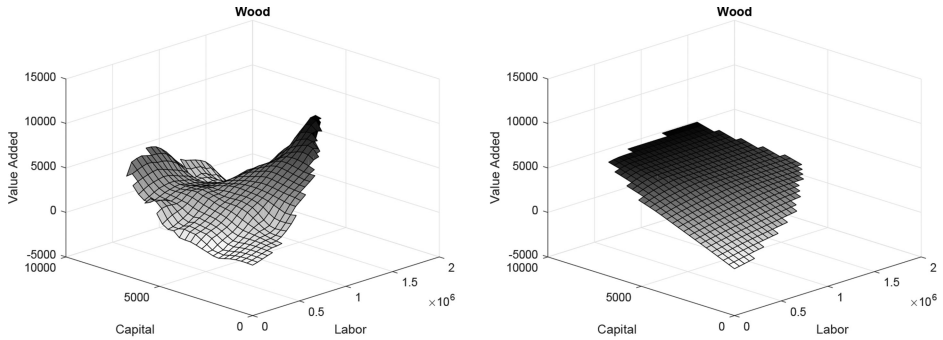
Table 8. Characteristics of the production function

	Plastic (2520)		
	Marginal productivity		
	Labor ($= b_l$) (million peso/man hours)	Capital ($= b_k$) (peso/peso)	Marginal rate of substitution ($= b_k/b_l$)
10th percentile	0.00396	0.111	23.3
25th percentile	0.00523	0.139	23.9
50th percentile	0.00579	0.139	24.0
75th percentile	0.00579	0.139	35.3
90th percentile	0.00579	0.260	44.8
	Wood (2010)		
	Marginal productivity		Marginal rate of substitution ($= b_k/b_l$)
	Labor ($= b_l$)	Capital ($= b_k$)	
10th percentile	1.46×10^{-18}	0.816	760
25th percentile	8.55×10^{-16}	0.816	760
50th percentile	0.00133	1.01	760
75th percentile	0.00133	1.01	9.73×10^{14}
90th percentile	0.00133	1.01	5.59×10^{17}



(a) Local Linear (b) SCKLS

Figure 3. Production function estimated by LL and SCKLS for the plastic industry (2520).



(a) Local Linear (b) SCKLS

Figure 4. Production function estimated by LL and SCKLS shape constraints for the wood industry (2010).

Table 9. Coefficient of contextual variables from a two-stage model

	Plastic (2520)		Wood (2010)	
	Dummy of exporting	Share of exportingin sales	Dummy of exporting	Share of exportingin sales
Point estimate	334.5	303.7	−763.0	4114
95% lower bound	148.7	−334.3	−1944	2568
95% upper bound	520.3	941.8	417.7	5660
p-value	4.70×10^{-4}	0.3493	0.2033	5.64×10^{-7}

substitution are also hard to interpret. In particular, it is not possible to identify most productive scale size.

Table 9 reports the estimated coefficients for the exporting variables. In the plastic industry, the dummy variable for exporting is significant and positive while exports' share of sales is not. This indicates that the plants that export tend to produce more output than plants that do not export regardless of the export quantity. In contrast, the coefficient on the exports' share of sales is significant and positive in the wood industry, while the dummy variable for exporting is not significant, indicating that establishments in the wood industry tend to be more productive the more they export. Thus, in both industries we find evidence of increased productivity for exporting firms.

Table 10 reports the most productive scale size for the 10, 25, 50, 75, 90 percentiles of Capital/Labor ratio distribution of observed input. In both industries, the observed value added output is the largest for establishments with high capital to labor ratios, indicating that capital-intensive establishments have increased actual output. Furthermore, labor-intensive establishments have smaller most productive scale size in both industries. This is consistent with the theory of the firm, that is, firms grow and become more capital intensive over time by automating processes with capital and using less labor.

Table 10. Most productive scale size for each capital/labor ratio

	Plastic (2520)		
Capital/Labor percentile	MPSS Labor	MPSS Capital	Output (value added)
10th percentile	619580	519.1	3290
25th percentile	529980	1344	3010
50th percentile	529980	2604	3185
75th percentile	529980	5617	3602
90th percentile	529980	10270	4248
	Wood (2010)		
Capital/Labor percentile	MPSS Labor	MPSS Capital	Output (value added)
10th percentile	2531100	741.6	1659
25th percentile	1045000	1200	2142
50th percentile	867250	2712	3470
75th percentile	662700	4179	4682
90th percentile	458150	5644	5893

7. CONCLUSION

This article proposed the SCKLS estimator that imposes shape constraints on a local polynomial estimator. We show the consistency and convergence rate of this new estimator under monotonicity and concavity constraints, as well as its relationship with CNLS and CWB. In applications where out-of-sample performance is less critical and the boundary behavior is of less concern, such as regulation applications, the CNLS estimator may be preferable because of its simplicity. In contrast, in cases where out-of-sample performance is important, such as survey data, the SCKLS estimator appears to be more robust. Simulation results reveal the SCKLS estimator outperforms CNLS and LL in most scenarios. We propose and validate the usefulness of several extensions, including variable bandwidth and non-uniform griding, which are important to estimate functions with non-uniform input dataset which is common in manufacturing survey and census data. We also propose a test for the imposed shape constraints based on SCKLS. Finally, we demonstrate the SCKLS estimator empirically using Chilean manufacturing data. We compute marginal productivity, marginal rate of substitution, most productive scale size and the effects of exporting, and provide several economic insights.

One limitation of the proposed SCKLS estimator is its computation efficiency due to the large number of constraints. The algorithm we proposed for reducing constraints performs well, and we demonstrate the ability to solve large problems instances within a reasonable time. Furthermore, our simulation results show good functional estimates even with a rough grid. Consequently, we can make use of the flexibility of the evaluation points to reduce the computational time of the estimator.

Potential future research could focus on the bandwidth selection methods. Typically, optimal bandwidth selection methods without shape constraints try to trade bias and variance to find the best estimator in terms of RMSE. Since the imposed shape restrictions already constrain the variance of the estimator to some extent, we expect that the optimal bandwidth in the SCKLS estimator will be smaller than the optimal unconstrained estimator. Further, if systematic inefficiency is present in the data, deconvoluting the residuals following the stochastic frontier literature would allow the investigation of a production frontier.

SUPPLEMENTARY MATERIALS

Appendix: The document contains: (A) extensions and the relationship between estimators; (B) technical proofs of the theoretical results; (C) a test of affinity using SCKLS; (D) an algorithm for SCKLS computational performance; (E) comprehensive results of existing and additional numerical experiments; (F) semiparametric model to integrate contextual variable; and (G) details of the application to the Chilean manufacturing data.

ACKNOWLEDGMENTS

We thank two anonymous reviewers and the Associate Editor for providing useful suggestions that helped improve this article. We also thank Chris Parmeter, Jeff Racine, and Qi Li for their helpful comments.

[Received May 2016. Revised January 2018.]

REFERENCES

Ackerberg, D. A., Caves, K., and Frazer, G. (2015), "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451. [8]

Alvarez, R., and Görg, H. (2009), "Multinationals and Plant Exit: Evidence From Chile," *International Review of Economics & Finance*, 18, 45–51. [8]

Andrews, D. W. K. (2000), "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405. [5]

Benavente, J. M. (2006), "The Role of Research and Innovation in Promoting Productivity in Chile," *Economics of Innovation and New Technology*, 15, 301–315. [8]

Beresteanu, A. (2005), "Nonparametric Analysis of Cost Complementarities in the Telecommunications Industry," *RAND Journal of Economics*, 36, 870–889. [9]

——— (2007), "Nonparametric Estimation of Regression Functions under Restrictions on Partial Derivatives," working paper. [1]

Bernard, A. B., and Jensen, J. B. (2004), "Exporting and Productivity in the USA," *Oxford Review of Economic Policy*, 20, 343–357. [8]

Birke, M., and Dette, H. (2007), "Estimating a Convex Function in Nonparametric Regression," *Scandinavian Journal of Statistics*, 34, 384–404. [1]

Brunk, H. D. (1955), "Maximum Likelihood Estimates of Monotone Parameters," *The Annals of Mathematical Statistics*, 26, 607–616. [1]

Carroll, R. J., Delaigle, A., and Hall, P. (2011), "Testing and Estimating Shape-Constrained Nonparametric Density and Regression in the Presence of Measurement Error," *Journal of the American Statistical Association*, 106, 191–202. [1]

Cavaliere, G., Nielsen, H.B., and Rahbek, A. (2017), "On the Consistency of Bootstrap Testing for a Parameter on the Boundary of the Parameter Space," *Journal of Time Series Analysis*, 38, 513–534. [5]

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836. [3]

Davidson, R., and Flachaire, E. (2008), "The Wild Bootstrap, Tamed at Last," *Journal of Econometrics*, 146, 162–169. [5]

De Loecker, J. (2007), "Do Exports Generate Higher Productivity? Evidence From Slovenia," *Journal of International Economics*, 73, 69–98. [8]

Du, P., Parmeter, C. F., and Racine, J. S. (2013), "Nonparametric Kernel Regression With Multiple Predictors and Multiple Shape Constraints," *Statistica Sinica*, 23, 1347–1371. [1]

Fan, Y., and Guerre, E. (2016), "Multivariate Local Polynomial Estimators: Uniform Boundary Properties and Asymptotic Linear Representation," in *Essays in Honor of Aman Ullah*, eds. G. Gonzalez-Rivera, R. C. Hill, and T.-H. Lee, Bingley, UK: Emerald, pp. 489–537. [3]

Grenander, U. (1956), "On the Theory of Mortality Measurement: Part II," *Scandinavian Actuarial Journal*, 1956, 125–153. [1]

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001), "Estimation of a Convex Function: Characterizations and Asymptotic Theory," *The Annals of Statistics*, 29, 1653–1698. [1]

Hall, P., and Heckman, N. E. (2000), "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *The Annals of Statistics*, 28, 20–39. [5]

Hall, P., and Huang, L.-S. (2001), "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics*, 29, 624–647. [1]

Hanson, D., and Pledger, G. (1976), "Consistency in Concave Regression," *The Annals of Statistics*, 4, 1038–1050. [1]

Henderson, D. J., and Parmeter, C. F. (2015), *Applied Nonparametric Econometrics*, Cambridge, UK: Cambridge University Press. [9]

Hildreth, C. (1954), "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619. [1]

Kuosmanen, T. (2008), "Representation Theorem for Convex Nonparametric Least Squares," *The Econometrics Journal*, 11, 308–325. [1,3]

Kuosmanen, T., and Kortelainen, M. (2012), "Stochastic Non-Smooth Envelopment of Data: Semi-Parametric Frontier Estimation Subject to Shape Constraints," *Journal of Productivity Analysis*, 38, 11–28. [6]

Lee, C.-Y., Johnson, A. L., Moreno-Centeno, E., and Kuosmanen, T. (2013), "A More Efficient Algorithm for Convex Nonparametric Least Squares," *European Journal of Operational Research*, 227, 391–400. [2,6]

Levinsohn, J., and Petrin, A. (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70, 317–341. [8]

Li, Q., and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, NJ: Princeton University Press. [3]

Li, Z., Liu, G., and Li, Q. (2017), "Nonparametric Knn estimation with monotone constraints," *Econometric Reviews*, 36, 988–1006. [1,3]

Lim, E., and Glynn, P. W. (2012), "Consistency of Multidimensional Convex Regression," *Operations Research*, 60, 196–208. [1,4]

Liu, R. Y. (1988), "Bootstrap Procedures Under Some Non-I.I.D. Models," *The Annals of Statistics*, 16, 1696–1708. [5]

Mammen, E. (1991), "Nonparametric Regression Under Qualitative Smoothness Assumptions," *The Annals of Statistics*, 19, 741–759. [1]

——— (1993), "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *The Annals of Statistics*, 21, 255–285. [5]

Masry, E. (1996), "Multivariate Local Polynomial Regression For Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571–599. [3]

Mazumder, R., Choudhury, A., Iyengar, G., and Sen, B. (2017), "A computational framework for multivariate convex regression and its variants," *Journal of the American Statistical Association*, (Accepted). [3]

Nesterov, Y. (2005), "Smooth Minimization of Non-Smooth Functions," *Mathematical Programming*, 103, 127–152. [3]

Pavcnik, N. (2002), "Trade Liberalization, Exit, and Productivity Improvements: Evidence From Chilean Plants," *The Review of Economic Studies*, 69, 245–276. [8]

Racine, J., and Li, Q. (2004), "Nonparametric Estimation of Regression Functions With Both Categorical and Continuous Data," *Journal of Econometrics*, 119, 99–130. [3]

Seijo, E., and Sen, B. (2011), "Nonparametric Least Squares Estimation of a Multivariate Convex Regression Function," *The Annals of Statistics*, 39, 1633–1657. [1]

Stone, C. J. (1977), "Consistent Nonparametric Regression," *The Annals of Statistics*, 5, 595–620. [3,6]

Varian, H. R. (1984), "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597. [2]

Wu, C.-F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics*, 14, 1261–1295. [5]