Nonparametric Estimation and Inference of Production Risk

Zheng Li, North Carolina State University, zli42@ncsu.edu
Roderick M. Rejesus, North Carolina State University, rod_rejesus@ncsu.edu
Xiaoyong Zheng, North Carolina State University, xzheng@ncsu.edu


Copyright 2018 by Zheng Li, Roderick M. Rejesus, Xiaoyong Zheng. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Nonparametric Estimation and Inference of Production Risk

Zheng Li∗ Roderick M. Rejesus† Xiaoyong Zheng‡

May 23, 2018

Abstract

This paper proposes a nonparametric approach for estimation of stochastic production functions with categorical variables, and then develops procedures that allow for inference on production risk. The estimation is based on the kernel method and the inference is based on a bootstrapping approach. We establish the asymptotic properties of our proposed estimator. Monte Carlo simulation results suggest that, compared with existing parametric methods, our proposed nonparametric procedure is more precise in estimation and more powerful in inference. In addition, we empirically illustrate the proposed nonparametric method using long-run corn production data from university field trials in Wisconsin that examines performance of genetically-modified (GM) varieties.

∗Department of Agricultural and Resource Economics, North Carolina State University, e-mail: zli42@ncsu.edu.
†Department of Agricultural and Resource Economics, North Carolina State University, e-mail: rod_rejesus@ncsu.edu.
‡Department of Agricultural and Resource Economics, North Carolina State University, e-mail: xzheng@ncsu.edu.
1 Introduction

Since the seminal work of Just and Pope (1978, 1979), numerous studies have employed the Just-Pope (JP) stochastic production function to estimate the effects of inputs on production/output and production/output risk.\(^1\) The JP production function has the general form of

\[
y = f(x) + \sqrt{h(x)} \varepsilon, \tag{1.1}
\]

where \(y\) is the output, \(x\) is the vector of inputs, \(f(\cdot)\) is the mean output function (or the deterministic part), \(h(\cdot)\) is the variance function (or the risk part) and \(\varepsilon\) is an exogenous production shock with \(E(\varepsilon|x) = 0\) and \(E(\varepsilon^2|x) = 1\). In most empirical studies, \(f(\cdot)\) and \(h(\cdot)\) are assumed to have specific functional forms and then (1.1) is estimated using either the feasible generalized least squares method (FGLS) or the maximum likelihood estimation method (MLE).\(^2\)

In particular, the FGLS estimation procedure consists of three steps. First, OLS or nonlinear least squares (depending on the specifications) is applied to (1.1) to obtain a consistent but inefficient estimate of \(f(\cdot)\). In the second step, squares of residuals (or functions of them) from the first step are regressed on the explanatory variables to obtain a consistent estimate of \(h(\cdot)\). Finally, in the third step, using the estimated \(\frac{1}{\sqrt{h(\cdot)}}\) from the second step as weights, the generalized least squares method is applied to (1.1) to obtain a consistent and efficient estimator of \(f(\cdot)\).

Recent studies have also extended the JP method along various dimensions, including: (a) examining the effects of inputs on higher moments of output, such as skewness and kurtosis, (Antle, 1983; Antle and Goodger, 1984) and on different quantiles of the output (Chavas and Shi, 2015), (b) joint estimation of production and utility functions (Love and Buccola, 1991), (c) correcting for endogeneity of certain inputs (Koundouri and Nauges, 2005), (d) examining and testing the asymmetric effects of inputs on the output (Antle, 2010), and (e) nonparametric estimation (Kumbhakar and Tsionas, 2010).

With the exception of Kumbhakar and Tsionas (2010), studies in this literature primarily employ the parametric approach for estimation. That is, particular functional forms are assumed for \(f(\cdot)\) and \(h(\cdot)\) and then estimation follows. As is well known, the traditional parametric methods suffer from the risk of mis-specification. When an econometric model is mis-specified, estimates will be inconsistent and hypothesis tests based on the estimates may have insufficient test power.

\(^{1}\) Examples include Hurd (1994), Traxler et al. (1995), Smale et al. (1998), Tveteros (1999), Di Falco et al. (2007), McCarl et al. (2008), to name just a few.

\(^{2}\) Estimating (1.1) using MLE also requires an additional assumption on the distribution of \(\varepsilon\). This assumption is not required with the FGLS method.
In this study, we propose a robust alternative estimation method to the traditional parametric approach. In particular, we propose a nonparametric approach to estimate the effects of inputs on mean output levels, as well as higher order moments of the output. Building upon recent advances in nonparametric econometrics by Li and Racine (e.g. Li and Racine, 2003; Racine and Li, 2004; Li and Racine, 2007), our proposed method also allows for both continuous and categorical input variables. Note that proper treatment of categorical variables in nonparametric estimation of JP models have not been fully explored in previous studies. In addition to the estimation method developed, we also propose a bootstrapping-based nonparametric procedure to conduct hypothesis testing in this framework. Overall, the proposed nonparametric method allows one to estimate production risk and conduct inferences without imposing strong parametric assumptions. Hence, results obtained from the proposed nonparametric procedures are more robust than those from the traditional parametric approach.

We then examine the performance of our new nonparametric estimator and testing procedure, and contrast them with those of the parametric approach, using Monte Carlo simulations. The simulation results show that our nonparametric estimation and testing procedures are superior relative to the parametric one when the functional forms are mis-specified. Finally, we empirically illustrate our proposed nonparametric estimation and testing approach by examining the effects of different genetically-modified (GM) corn varieties on mean production and production risk using long-run university field trial data from Wisconsin.

The closest paper to ours in spirit is Kumbhakar and Tsionas (2010). Kumbhakar and Tsionas (2010) also propose a nonparametric method to estimate the impacts of inputs on the mean and variance (risk) of output. Our study here is different from theirs in three important aspects. First, our method allows for categorical or discrete inputs while Kumbhakar and Tsionas (2010) only consider inputs that are continuous. Many input variables such as the type of seeds (or crop varieties) used, location dummies, and choice of tillage practices (e.g., no-till versus conventional till), are categorical rather than continuous variables. Thus, our method significantly expands the range of potential applications of the nonparametric method for production function and risk estimation. Second, while Kumbhakar and Tsionas (2010) focus on nonparametric estimation only, we propose both a nonparametric estimation and a nonparametric hypothesis testing procedure. An inference procedure is necessary to answer various economic questions such as whether one type of crop variety is risk-reducing compared with another variety (e.g., traditional vs. GM crops). Third, while Kumbhakar and Tsionas (2010) focus on estimating variance as a measure of risk, we estimate not only the variance, but higher moments of the output as well, such as skewness and kurtosis. Higher moments offer a more complete picture of the effects of inputs on production risk.
Also related to the current paper are studies estimating the densities of the output or yields using nonparametric and semiparametric methods (e.g. Goodwin and Ker, 1998; Ker and Goodwin, 2000; Ker and Coble, 2003). Different from our study here, these papers do not explicitly examine the impacts of inputs on output levels and risk, but rather on providing a better characterization of the yield distribution for insurance rating purposes.

The rest of the paper is organized as follows. Section 2 presents our nonparametric estimation and inference procedures and contrast them with those of a traditional parametric approach. In Section 3, we conduct Monte Carlo simulations to demonstrate the performance of our estimator, and the corresponding testing procedures. In Section 4, we apply our method to study the impacts of GM corn varieties on production/yield and production/yield risk. The final section concludes.

2 Estimation and Inference

Consider a stochastic production function,

\[ y = g(x, z) + u, \]

where \( y \) is the output, \( x = \{x_1, x_2, \cdots, x_{d_1}\} \in D_x \) is a \( d_1 \times 1 \) vector of continuous explanatory variables (e.g. fertilizer, planting density), \( z = \{z_1, z_2, \cdots, z_{d_2}\} \in D_z \) is a \( d_2 \times 1 \) vector of categorical explanatory variables (e.g. location, seed type), and \( u \) is a random shock (e.g. unobserved weather, plant diseases, pests) with \( E(u|x, z) = 0 \). Although the first moment of \( u \) conditional on the explanatory variables is assumed to be 0 for identification purpose, our model imposes no restriction on the higher order moments of \( u \) conditional on the explanatory variables.

We are interested in estimating the mean, variance, skewness and kurtosis of \( y \) conditional on \( x \) and \( z \), i.e.,

\[ M_1(x, z) \equiv E(y|x, z) \]

and

\[ M_j(x, z) \equiv E\{[y - M_1(x, z)]^j|x, z\}, \quad j \geq 2. \]

We consider two methods: (i) a parametric regression approach, and (ii) a nonparametric regression procedure. The parametric regression assumes a specific functional form for the regressors, while the nonparametric regression does not assume any specific functional form for the regressors. The details of the two methods are as follows.
2.1 Nonparametric Estimation

In this subsection, we propose to use the kernel method to estimate the conditional moments. Let \( \{y_i, x_i, z_i\}_{i=1}^{n} \) be an independent and identically distributed \((i.i.d.)\) sample from the population and \( n \) the sample size. The kernel estimator for the mean of \( y \) conditional on \( x = x_0 \) and \( z = z_0 \) is

\[
\hat{M}_1(x_0, z_0) = \frac{\sum_{j=1}^{n} y_j K_h(x_j - x_0) \Lambda_\lambda(z_j, z_0)}{\sum_{j=1}^{n} K_h(x_j - x_0) \Lambda_\lambda(z_j, z_0)}
\]  

(2.1)

where \( K_h(\cdot) \) is the kernel function for continuous variables and \( \Lambda_\lambda(\cdot) \) is the smoothing function for categorical variables (Li and Racine, 2003; Racine and Li, 2004). The kernel function for continuous variables is,

\[
K_h(x_i - x_0) = \prod_{s=1}^{d_1} K(\frac{x_{i,s} - x_{0,s}}{h_s}),
\]

where \( K(\cdot) \) is the second order univariate kernel function. Let the kernel function \( K(\cdot) \) be a bounded, continuous and symmetric probability density function. Many candidates can be used as the univariate kernel function. Popular kernel functions include the Gaussian kernel and the Epanechnikov kernel. The Gaussian kernel is given by

\[
K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2},
\]

with support of the real line, and the Epanechnikov kernel is given by

\[
K(u) = \frac{3}{4}(1 - u^2),
\]

with support \(|u| \leq 1 \). The bandwidth for the \( s^{th} \) continuous variable is \( h_s > 0 \). Let the collection of bandwidths for continuous variables be \( h = (h_1, \ldots, h_{d_1})' \), and \( H = \prod_{s=1}^{d_1} h_s \).

For categorical variables, the smoothing function is,

\[
\Lambda_\lambda(z_i, z_0) = \prod_{s=1}^{d_2} l(z_{i,s}, z_{0,s}, \lambda_s)
\]

where if the categorical variable is un-ordered,

\[
l(z_{i,s}, z_{0,s}, \lambda_s) = \lambda_s^{I(z_{i,s} \neq z_{0,s})},
\]
where $I(\cdot)$ is the indicator function. If the categorical variable is ordered, the smoothing function is,

$$I(z_{i,s}, z_{0,s}, \lambda_s) = \lambda_s^{\mid z_{i,s} - z_{0,s} \mid}.$$ 

The tuning parameters for the $s^{th}$ categorical variable is $\lambda_s \in [0, 1]$. Let the collection of tuning parameters for categorical variables be $\lambda = (\lambda_1, \cdots, \lambda_{d_2})'$. Note that $\lambda_s = 0$ (assuming $0^0 = 1$) leads to subsampling based on the values of the $s^{th}$ categorical variable, and $\lambda_s = 1$ leads to pooling over the $s^{th}$ dimension of $z$. The conditional mean estimator in (2.1) is consistent and asymptotically normal, which is well established in Racine and Li (2004).

Next, we focus on estimating the conditional higher order moments. Given the estimated conditional mean $\hat{M}_1(x_0, z_0)$, we obtain the regression residual,

$$\hat{u}_i = y_i - \hat{M}_1(x_i, z_i).$$

The variance of $y$ conditional on $x = x_0$ and $z = z_0$ can be nonparametrically estimated by

$$\hat{M}_2(x_0, z_0) = \frac{\sum^n_{j=1} \hat{u}_j^2 K_h(x_j - x_0)\Lambda_{\lambda}(z_j, z_0)}{\sum^n_{j=1} K_h(x_j - x_0)\Lambda_{\lambda}(z_j, z_0)},$$

where the smoothing functions are defined in the same way as in (2.1). Similarly, the higher order moments of $y$ (e.g. skewness and kurtosis) conditional on $x = x_0$ and $z = z_0$ can be estimated by

$$\hat{M}_j(x_0, z_0) = \frac{\sum^n_{i=1} \hat{u}_i^j K_h(x_i - x_0)\Lambda_{\lambda}(z_i, z_0)}{\sum^n_{i=1} K_h(x_i - x_0)\Lambda_{\lambda}(z_i, z_0)}, \ j \geq 3.$$

To establish the asymptotic properties of our estimator, we first impose some assumptions and introduce some notations.

**Assumption 2.1.** The production function $g(\cdot, z)$ is twice continuously differentiable on $D_x$.

**Assumption 2.2.** The joint probability density function of $X_i$ and $Z_i$, $f(x, z)$, is bounded away from infinity and zero for $x \in D_x$ and $z \in D_z$.

**Assumption 2.3.** As $n \to \infty$, $h_s \to 0$ for $s = 1, 2, \cdots, d_1$, $nH \to \infty$ and $\lambda_s \to 0$ for $s = 1, 2, \cdots, d_2$.

**Assumption 2.4.** When estimating the $j^{th}$ conditional moment, we assume that the $(2j)^{th}$ conditional moment of $u_i$ exists given any $x \in D_x$ and $z \in D_z$, and the $j^{th}$ conditional moment $M_j(\cdot, z)$ is twice continuously differentiable on $D_x$.

Let $f_s^*(\cdot)$ be the first derivative of $f(\cdot)$ with respect to the $s$-th continuous variable. Let $M_2^s(\cdot)$ and $M_2^{ss}(\cdot)$ be the first and second derivatives of $M_2(\cdot)$ with respect to the $s$-th
continuous variable respectively. Let

\[ B_j(x, z) = \frac{\mu_2}{2f(x, z)} \sum_{s=1}^{d_1} h_s^2 [2f^s(x, z)M_j^s(x, z) + f(x, z)M_j^{ss}(x, z)] + \]

\[ \frac{1}{f(x, z)} \sum_{z' \in \mathcal{D}_z} \lambda_s I_s(z', z) (M_j(x, z') - M_j(x, z)) f(x, z') \]  (2.2)

and

\[ \tilde{B}_j(x, z) = jM_{j-1}(x, z)B_1(x, z) + B_j(x, z) \]  (2.3)

with \( \mu_2 = \int v^2 K(v) dv \) and

\[ I_s(z', z) = I(z' \neq z_s) \prod_{t=1, t \neq s}^{d_2} I(z_t = z_t). \]

Let

\[ \Omega_j(x, z) = \frac{\kappa^{d_1}}{f(x, z)} [M_2(x, z) - M_j^2(x, z)] \]  (2.4)

and

\[ \tilde{\Omega}_j(x, z) = \frac{\kappa^{d_1}}{f(x, z)} [j^2 M_2(x, z)M_{j-1}^2(x, z) - 2jM_{j-1}(x, z)M_{j+1}(x, z)] + \Omega_j(x, z), \]  (2.5)

with \( \kappa = \int K(v)^2 dv \).

**Proposition 2.1.** Under Assumptions 2.1-2.4, the higher order moment estimators \( \hat{M}_j(x, z) \), \( j \geq 2 \), are consistent and asymptotically normal, with

\[
\begin{cases}
\sqrt{nH}(\hat{M}_j(x, z) - M_j(x, z) - B_j(x, z)) \overset{d}{\rightarrow} N(0, \Omega_j(x, z)), & j = 2 \\
\sqrt{nH}(\hat{M}_j(x, z) - M_j(x, z) - \tilde{B}_j(x, z)) \overset{d}{\rightarrow} N(0, \tilde{\Omega}_j(x, z)), & j \geq 3
\end{cases}
\]

**Proof.** See Appendix. \( \square \)

Proposition 2.1 shows that: (1) the convergence rate is \( \sqrt{nH} \) for all of the conditional variance and higher order moment estimators, and (2) the convergence rate does not depend on the dimension of the categorical variables. If one deals with a dataset with a small number of continuous variables and some categorical variables, the convergence rate is satisfactory.

To implement the nonparametric estimation procedure outlined above, the tuning parameters, \( h \) and \( \lambda \), need to be chosen. The prevalent methods of selecting tuning parameters include the rule-of-thumb method and the least-squares cross-validation (CV) method (Racine and Li, 2004). Both methods have their own advantages and disadvantages. With
the least-squares CV method, one chooses $h$ and $\lambda$ by solving the following minimization problem,
\[
\min_{h,\lambda} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - \hat{M}_{1,(-i)}(x_i, z_i]\right]^2,
\]
where
\[
\hat{M}_{1,(-i)}(x_i, z_i) = \frac{\sum_{j=1, j\neq i}^{n} y_j K_h(x_j - x_0) \Lambda_{\lambda}(z_j, z_0)}{\sum_{j=1, j\neq i}^{n} K_h(x_j - x_0) \Lambda_{\lambda}(z_j, z_0)}
\]
is the leave-one-out estimator with the $i^{th}$ data point omitted. The least-squares CV is a data-driven method. It can be shown that the CV-selected bandwidths are asymptotically optimal (Racine and Li, 2004). A feature of this method is that when a regressor (continuous or categorical) is irrelevant to the dependent variable, the CV method selects a tuning parameter that smooths out the irrelevant regressor (Hall et al. (2007)). To be specific, the tuning parameter $h_s \to \infty$ smooths out continuous variables $x_s$, and $\lambda_s \to 1$ smooths out categorical variables $z_s$.

In empirical applications, one usually does not know the true relationship between the regressors and the dependent variable. Practitioners can then use the CV method to reduce the negative impact of redundant regressors. The disadvantage of the CV method is that when the dimension of regressors or sample size becomes large, the minimization of (2.6) could be difficult in terms of computation. By contrast, the rule-of-thumb method is computationally easy. The rule-of-thumb method provides guidelines for choosing tuning parameters for continuous variables. It suggests that one chooses tuning parameters $h_s = c_s \sigma_s n^{-1/(4+d_1)}$, where $\sigma_s$ is the standard deviation of the $s^{th}$ continuous variable and $c_s$ is a constant, which is often chosen to be close to one.³ After $h$ is chosen by the rule-of-thumb method, one can use the cross-validation method to choose $\lambda$. This method yields sub-optimal bandwidths with reduced computational burden. An iteration of CV selecting $h$ and $\lambda$ separately restores the optimality, as shown in Li and Zhou (2005).

2.2 Nonparametric Inference

Estimation is often just the first step in economic analysis and answering many economic questions of interest requires testing hypotheses using the estimates after estimation. For example, one may be interested in testing whether a new seed variety is yield increasing and/or risk reducing compared with a traditional variety, or one may be interested in testing whether increasing/decreasing the amount of an input has positive/negative impacts on the yield and risk. Therefore, we also propose a bootstrapping-based hypothesis testing

³The normal reference rule for density estimation suggest $c_s \approx 1.06$. 
procedures that one can use for such purposes. We consider two testing procedures that are suitable for continuous regressors and categorical regressors respectively.

2.2.1 Inference on Continuous Variables

Suppose we are interested in the effect of the first continuous regressor $x_1$ on the conditional moments. Let $x = \{x_1, \bar{x}\}$, with $\bar{x} = \{x_2, x_3, \ldots, x_{d_1}\}$. We want to test if the increase in $x_1$ is associated with the increase in the conditional moment, when $x_1 \in D$, where $D$ is the interval of $x_1$ that we want to examine. Since testing for increasing and decreasing are similar, we only consider testing for increasing. If the researcher has control over the rest of the inputs (e.g. fertilizer), one can specify the null hypothesis at fixed values of $\tilde{x}_0$ and $z_0$, i.e.,

$$H_0 : \frac{\partial M_j(x_1, \tilde{x}_0, z_0)}{\partial x_1} \geq 0, \forall x_1 \in D.$$  

If the inputs are random (e.g., rainfall), one can first take expectation over the random components. We define the integrated conditional moment as

$$IM_j(x_1) = \int \sum_z M_j(x_1, \tilde{x}, z)dW(\tilde{x}, z),$$  \hspace{1cm} (2.7)

where $W(\tilde{x}, z)$ is the joint CDF of $\tilde{x}$ and $z$, which can be easily estimated by the empirical CDF estimator, i.e.,

$$\hat{W}(\tilde{x}_0, z_0) = \frac{1}{n} \sum_{i=1}^{n} I(\tilde{x}_i \leq \tilde{x}_0, z_i \leq z_0).$$

Once we plug $\hat{W}(\tilde{x}_0, z_0)$ into (2.7), the integrated moment is in the form of sample mean,

$$IM_j(x_1) = \frac{1}{n} \sum_{i=1}^{n} M_j(x_1, \tilde{x}_i, z_i).$$

The corresponding null hypothesis is

$$H_0 : \frac{\partial IM_j(x_1)}{\partial x_1} \geq 0, \forall x_1 \in D,$$

and alternative hypothesis is

$$H_0 : \frac{\partial IM_j(x_1)}{\partial x_1} < 0, \forall x_1 \in D.$$  

If one encounters mixed random and non-random inputs, one can integrate out the selected (random) regressors. Our testing procedure is unified for no matter point-estimated moments
or integrated moments. We only present integrated moments in the rest of this section.

The monotonicity of a function is an infinite-dimensional problem. We first reduce it to a finite-dimensional problem. Consider a finite set of ordered points on $D$, i.e., $x_{11} < x_{12} < \cdots < x_{1N}$. The null hypothesis becomes

$$IM_j(x_{11}) \leq IM_j(x_{12}) \leq \cdots \leq IM_j(x_{1N}).$$

The researcher can select the number of points $N$ to achieve the desired precision. In the case when $j = 1$ (conditional mean regression), the monotonicity test can be performed following Hall and Huang (2001) or Du et al. (2013). However, those methods are not sufficient when $j > 1$, due to the complexity of the two-stage estimation procedure. Fang and Santos (2017) propose a new framework for testing monotonicity, which enables us to handle the cases with any order conditional moment. Define

$$\phi(IM_j) = \min_{\{y_i\}_{i=1}^N} \sum_{i=1}^N (IM_j(x_{1i}) - y_i)^2$$

s.t. $y_1 \leq y_2 \leq \cdots \leq y_N$.

$\{y_i\}_{i=1}^N$ is a sequence of increasing values. $\phi(IM_j)$ measures the distance between the function $IM_j$ and the closest increasing function. It is our test statistic. If $IM_j$ is an increasing function, then $\phi(IM_j) = 0$. The sample analogue $\phi(\hat{IM}_j)$ serves as the test statistic. If $\phi(\hat{IM}_j)$ is far enough from zero, we tend to reject the null hypothesis. Otherwise, we do not reject the null hypothesis.

The last step is to determine the critical values. Using the recent advancements in Fang and Santos (2017) and Hong and Li (2017). The critical values can be obtained by a delta-method-based bootstrap. The bootstrap procedure is as follows.

1. Re-sample the triplet $\{y_i, x_i, z_i\}_{i=1}^n$ with replacement to obtain a new sample $\{y_i^*, x_i^*, z_i^*\}_{i=1}^n$.

2. Estimate the integrated conditional moment $IM_j(x_1)$ from the bootstrap sample. Denote the estimates as $\hat{IM}^*(x_1)$. Let

$$h^* \equiv \hat{IM}^*(x_1) - \hat{IM}(x_1)$$

Obtain $\phi'_n(h^*)$ by

$$\phi'_n(h^*) = \frac{\phi(\hat{IM}(x_1) + t_n \sqrt{nH} h^*) - \phi(\hat{IM}(x_1))}{t_n \sqrt{nH}}$$
(3) Repeat step (1) and step (2) for $B$ times, and collect all $B$ values of $\phi_n'(h^*)$. The upper $\alpha$-th quantile of $\phi_n'(h^*)$ is the critical value for significance level $\alpha$.

There is a tuning parameter in step (2). Hong and Li (2017) suggest that $t_n$ goes to 0 at the rate of $n^{-\frac{1}{3}}$.

### 2.2.2 Inference on Categorical Variables

Suppose we are interested in the effect of the first categorical regressor (e.g. the indicator of seed varieties), we let $z = \{z_1, \tilde{z}\}$, with $\tilde{z} = \{z_2, z_3, \cdots, z_d\}$. Suppose $z_1$ takes $m$ values, that is, $z_1 \in \{1, 2, \cdots, m\}$. Define the integrated conditional moment for variety $k$ as

$$IM_j(k) = \int \sum_{\tilde{z}} M_j(x, k, \tilde{z})dW(x, \tilde{z}),$$

where $W(x, \tilde{z})$ is the joint CDF of $x$ and $\tilde{z}$. Define that seed variety $k$ dominates variety $k'$ in $j^{th}$ order conditional moment if $IM_j(k) > IM_j(k')$ and let the distance between the integrated conditional moments of $k$ and $k'$ be

$$D(k, k') = IM_j(k) - IM_j(k').$$

Our goal here is to determine if there is a statistically significant difference in conditional moments between the two different seed varieties. Therefore, the null hypothesis is,

$$H_0 : D(k, k') = 0$$

and the alternative hypothesis is,

$$H_1 : D(k, k') \neq 0.$$

The test statistic $D(k, k')$ is estimated by

$$\hat{D}(k, k') = \hat{IM}_j(k) - \hat{IM}_j(k') = \int \sum_{\tilde{z}} \hat{M}_j(x, k, \tilde{z}) - \hat{M}_j(x, k', \tilde{z})dW(x, \tilde{z}),$$

with the bandwidth of $z_1$ in $\hat{M}_j(x, z_1, \tilde{z})$ being zero. One can also specify a one-sided test such as $H_0 : D(k, k') \geq 0$ versus $H_1 : D(k, k') < 0$. The testing procedure is similar to the two-sided test.
We use re-sampling method to construct confidence intervals for the two-sided test statistic \( \hat{D}(k, k') \). For example, for a level \( c \in (0, 1) \) test, we use the re-sampling bootstrap method to obtain a \( (1 - c)\% \) confidence interval for \( \hat{D}(k, k') \), say, \([Q_{c/2}, Q_{1-c/2}]\). We reject \( H_0 \) if 0 is outside of \([Q_{c/2}, Q_{1-c/2}]\) and we do not reject \( H_0 \) otherwise. To determine the critical values needed for the test, we use the following bootstrapping procedure to estimate the distribution of the test statistic \( D(k, k') \):

1. Re-sample the triplet \( \{y_i, x_i, z_i\}_{i=1}^n \) with replacement to obtain a new sample \( \{y_i^*, x_i^*, z_i^*\}_{i=1}^n \).

2. Estimate the model and obtain the test statistic \( TS^* (= \hat{D}(k, k') \text{ in this example}) \) based on the new sample generated in step (1).

3. Repeat step (1) and (2) for \( B \) times to obtain a set of test statistics \( \{TS^*_i\}_{i=1}^B \).

4. Set the significance level as \( c \). Obtain the \( c/2 \) and \( 1 - c/2 \) quantiles of \( \{TS^*_i\}_{i=1}^B \), denoted as \( Q_{c/2} \) and \( Q_{1-c/2} \), respectively.

5. If \( 0 \in [Q_{c/2}, Q_{1-c/2}] \), then we cannot reject the null hypothesis at \( c \) significance level. Otherwise, we reject the null hypothesis at \( c \) significance level.

The validity and statistical properties of this bootstrapping based testing procedure can be established through standard arguments (e.g. Hall, 1992).

2.3 Parametric Method

For comparison purpose, in this subsection, we briefly describe the production function estimation and testing procedures within the traditional parametric framework. When both \( f(\cdot) \) and \( h(\cdot) \) are specified to be linear (both in parameters and explanatory variables), estimation proceeds as follows (see e.g. Antle, 1983). In the first stage, OLS is applied to estimate the following equation,

\[
y_i = \alpha_1 + x_i'\beta_1 + z_i'\gamma_1 + v_i
\]

where \( \alpha_1, \beta_1 \) and \( \gamma_1 \) are parameters. The regressor \( z \) contains dummy variables for all but one possible values of each categorical variable. The residuals are obtained using

\[
\hat{v}_i = y_i - \hat{\alpha}_1 - x_i'\hat{\beta}_1 - z_i'\hat{\gamma}_1,
\]

where \( \hat{\alpha}_1, \hat{\beta}_1 \) and \( \hat{\gamma}_1 \) are the OLS estimates. In the second-stage, the following regressions are estimated,

\[
\hat{v}_i^j = \alpha_j + x_i'\beta_j + z_i'\gamma_j + e_{j,i}, \quad j = 2, 3, 4.
\]
where $\alpha_j$, $\beta_j$ and $\gamma_j$ ($j = 2, 3, 4$) are parameters in the $j$th regression to examine how inputs impact the higher moments of output. To test the same hypothesis as in Section 2.2, we test if the corresponding coefficient is significant different from zero. Note that standard statistical inference for linear regression model is not correct in the second stage. One could use a similar bootstrap procedure as in Section 2.2 to obtain the correct statistical inference.

3 Monte Carlo Studies

Before turning to the empirical application, we first conduct several Monte Carlo experiments to assess the performance of our nonparametric estimator and testing procedure, and compare these with those of the parametric approach. We use the Integrated Mean Squared Error (IMSE) as an estimation performance measure, which is obtained by

$$IMSE_j = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (\hat{M}_{j,k}(x_i, z_i) - M_j(x_i, z_i))^2, \quad j = 1, 2, 3, 4.$$  

where $m$ is the number of replications, and $\hat{M}_{j,k}(x_i, z_i)$ is the value of $\hat{M}_j(x_i, z_i)$ in $k^{th}$ simulation.

We use the rejection rates as our performance measures in testing. When the null hypothesis is true, the rejection rates approximate the test size, and when the null hypothesis is false, the rejection rates approximate the test power. A test with good performance is defined by a combination of correct test size and strong test power.

3.1 Performance in Estimation

The data generating process (DGP) for our first experiment is as follows. Consider a stochastic production function

$$y = 2\sin(\pi x) + (1 + x) \cdot u,$$

where $x \sim \text{Uniform}[0, 1]$ and $u$ follows a mixed normal distribution with $\frac{1}{4}N(-\frac{3}{2}, \frac{1}{4}) + \frac{3}{4}N(\frac{1}{2}, \frac{1}{4})$. It is easy to see that $u$ has zero mean, unit variance and non-zero skewness, which is helpful in examining the estimation and inference on the conditional skewness estimator. The theoretical moments of $y$ conditional on $x$ are

$$M_1(x) = 2\sin(\pi x), \quad M_2(x) = (1 + x)^2,$$

$$M_3(x) = -(1 + x)^3 \cdot \left(\frac{3}{4}\right), \quad M_4(x) = (1 + x)^4 \cdot \frac{21}{8}.$$
The kernel function we use in the nonparametric method is Gaussian, and the bandwidth is selected by rule-of-thumb. The sample sizes are 100, 200 and 400. The number of replications is 1000. The number of bootstrap is 200.

We first look at the IMSEs, which are reported in Table 1. The method “P” stands for the parametric method, and “NP” stands for the nonparametric method. Results show that the behavior of the nonparametric method is normal. As the sample size increases, the IMSE decreases at approximately the rate of $\sqrt{nh}$. The IMSEs for the parametric method decreases very slow as the sample size increases. This is because of the inconsistency of the parametric method. Since the mean function of the DGP is nonlinear, the parametric method is not consistent in estimating the conditional mean. As a result, in the second stage, one cannot obtain consistent estimates for higher order conditional moments. To further illustrate the inconsistency of the parametric method, and the consistency of the nonparametric method, we examine the two components (squared bias and variance) of IMSEs. The integrated variance (IVAR) and integrated squared bias (IBIAS) are obtained by

$$ IVAR_j = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (\hat{M}_{j,k}(x_i, z_i) - \bar{M}_j(x_i, z_i))^2, \quad j = 1, 2, 3, 4, $$

where $\bar{M}_j(x_i, z_i) = \frac{1}{m} \sum_{k=1}^{m} \hat{M}_{j,k}(x_i, z_i)$, and

$$ IBIAS_j = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} (\bar{M}_j(x_i, z_i) - M_j(x_i, z_i))^2, \quad j = 1, 2, 3, 4. $$

We report IBIAS and IVAR in Tables 2 and 3. The comparison in IBIAS indicates that the nonparametric estimator is asymptotically unbiased in all cases. For the parametric estimator, the squared bias is stable (does not shrink to zero) as the sample size increases. The comparison in IVAR shows that both estimators have variances shrinking to zero as the sample size increases. This indicates that even though the parametric estimator is not consistent (asymptotically biased), it is still convergent in probability.

<table>
<thead>
<tr>
<th>Method</th>
<th>P</th>
<th>NP</th>
<th>P</th>
<th>NP</th>
<th>P</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=100</td>
<td></td>
<td>n=200</td>
<td></td>
<td>n=400</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>42.41</td>
<td>10.35</td>
<td>40.21</td>
<td>6.27</td>
<td>39.07</td>
<td>3.82</td>
</tr>
<tr>
<td>Variance</td>
<td>40.31</td>
<td>31.37</td>
<td>28.37</td>
<td>17.93</td>
<td>21.69</td>
<td>10.60</td>
</tr>
<tr>
<td>Skewness</td>
<td>491.61</td>
<td>257.73</td>
<td>349.41</td>
<td>149.19</td>
<td>279.80</td>
<td>92.03</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>18911.81</td>
<td>6243.66</td>
<td>13989.08</td>
<td>3402.98</td>
<td>11066.14</td>
<td>2017.84</td>
</tr>
</tbody>
</table>

Table 1: Comparison in Integrated Mean Squared Errors ($\times 10^{-2}$)
### Performance in Testing

To assess the performance of our testing procedure, we run a second Monte Carlo experiment. Consider testing if the conditional moments are increasing/decreasing in $x$ on $[0, 1]$. It is easy to see that the conditional mean is neither increasing nor decreasing in $x$. The conditional variance and kurtosis are increasing in $x$. The conditional skewness is decreasing in $x$.

Tables 4 and 5 report the rejection rates in testing increasing and decreasing respectively. In both tables, we label $H_0$ and $H_1$ indicating the null hypothesis being true and false respectively. The significance level is 5%. In the case of conditional mean, the parametric method has low rejection rates in both tests (increasing and decreasing). This is because the parametric method uses a flat curve to approximate the sine function in the DGP. On the other hand, from the rejection rates of the nonparametric method, we see that the nonparametric method can properly detect that the sine function is neither increasing nor decreasing in $x$ on $[0, 1]$. For higher order conditional moments, both tests behave well.

When the null hypothesis is true, the rejection rates are under the nominal test size (second and fourth rows in Table 4, and third row in Table 5). Since the monotonicity restriction in the null hypothesis is not binding, i.e., the inequality between the first derivative and zero holds instead of equality, the rejection rates go to zero as the sample size increases. When the null hypothesis is false, the rejection rates go to one as the sample size increases (third row in Table 4, and second and fourth rows in Table 5). Due to the inaccurate first stage estimation, the parametric method tend to have lower test power compared with the nonparametric method.
We then consider the test on the categorical variables. We add a categorical component into the DGP, which gives

\[ y = 2\sin(\pi x) + \frac{z}{2} + (1 + x + \frac{z}{2}) \cdot u, \]

where \( x \) and \( u \) follow the same distributions as in the previous case, and \( z \sim \text{Binomial}(1, 0.5) \).

It is easy to see that \( z \) plays a role in all of the four conditional moments. We test if the effect of \( z \) is significant different from zero in all of the four conditional moments using both parametric and nonparametric methods. The rejection rates are reported in Table 6. The significance level is 5%. In all of the cases, the rejection rates go to one as the sample size increases. The parametric method suffers from the loss of test power due to the inaccurate first stage estimation. The test power of the nonparametric method is satisfactory.
4 Empirical Application

In this section, we empirically illustrate the usefulness of our proposed nonparametric estimation and testing procedures by examining the production risk effects of GM corn varieties based on long-run university field trial data.

4.1 Data Description

Our empirical application is based on plot-level data from field experiments conducted by University of Wisconsin researchers for the period 1990 to 2010. The primary purpose of these trials was to examine the yield performance of a number of different corn varieties (or hybrids), specifically GM versus non-GM corn varieties (i.e., especially since GM corn became commercially available in 1996). The field experiments were conducted at agricultural research stations and with long-term farmer cooperators primarily in 12 sites in Wisconsin. Each site has approximately 800 to 1,000 plots. The experimental design for the trials was a randomized complete block in which each variety (or hybrid) is grown in at least three separate, randomly-assigned plots at each site to account for yield variability. Management practices for these sites were typical of those used on corn farms practicing rainfed agriculture in the Midwest. Relevant data are then collected (e.g., yields, fertilizer application, etc.) in each plot. The full data set has 31,799 observations, with each observation associated with a single variety at a single location for a single year (See the top panel of Table 7 for the summary statistics associated with the full data set).

Our analysis here relies on a sub-set of the full 31,799 observation data. Specifically, we only focus on the production risk effects of corn varieties with the following traits: (a) those with a single trait that only confers insect resistance, and (b) those with “double-stacked” traits that confers both insect resistance and herbicide tolerance. The single-trait we chose for our analysis is the one that provides resistance to the European Corn Borer (henceforth, the ECB). This was one of the first GM trait made available to farmers that had demonstrated market success. The other kind of GM variety we examine is the “double-stack” type where it has resistance to ECB and tolerance to a particular herbicide called glufosinate (hereinafter, we call this double-stack variety ECB-GFT). The ECB-GFT variety is one of the more popular double-stack GM varieties used by farmers when these types of two-trait varieties

---

4The data set used in this study is the same one utilized in Shi et al. (2013). We thank Guanming Shi at the University of Wisconsin for graciously sharing the data and their estimation code with us. More detailed information about the data and the field experiments can be seen in p. 112 of Shi et al. (2013).

5From Shi et al. (2013), about 4,748 varieties were tested over the data period, where approximately half were conventional non-GM varieties and the other half were GM varieties of some kind. The data contains information on 12 kinds of GM varieties with different traits and trait combinations.
were initially released. With our focus on measuring the risk effects of GM varieties with ECB and ECB-GFT (vis-a-vis the conventional variety), the final data set we use for our empirical analysis has 24,134 plot-level observations (See the bottom panel of Table 7 for the summary statistics associated with the data set used in our empirical application below). There are 19,652 observations for conventional variety, 3,484 observations for ECB, and 998 observations for ECB-GFT.

The dependent variable in our empirical specification is a continuous yield variable (in bushels per acre). The explanatory variables of main interest are the ECB and ECB-GFT dummy variables. These two categorical variables allow us to show the usefulness of the proposed nonparametric approach in terms of estimating stochastic production functions with categorical variables, and providing inference related to the risk effects of these two particular categorical variables. The other plot-level control variables included in our specification are the following: planting density (in 1000s of plants per acre), amount of fertilizer applied (in lbs per acre), whether or not there is irrigation (dummy variable equal to 1 if there is irrigation, zero otherwise), whether or not insecticide was applied (dummy variable equal to 1 if insecticide was applied, zero otherwise), year dummy variables, and location dummy variables. In summary, the empirical specification has two continuous explanatory variables (e.g., planting density and fertilizer application) and the rest of the explanatory variables are categorical.

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (bushels per acre)</td>
<td>31,799</td>
<td>178.62</td>
<td>40.04</td>
<td>21.00</td>
<td>289.81</td>
</tr>
<tr>
<td>Planting Density (1000s per acre)</td>
<td>31,799</td>
<td>28.49</td>
<td>1.93</td>
<td>18.25</td>
<td>33.41</td>
</tr>
<tr>
<td>Fertilizer (lb per acre)</td>
<td>31,799</td>
<td>130.17</td>
<td>47.31</td>
<td>0.50</td>
<td>236.25</td>
</tr>
<tr>
<td>Irrigation</td>
<td>31,799</td>
<td>0.10</td>
<td>0.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Insecticide</td>
<td>31,799</td>
<td>0.35</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Subsample for Estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield (bushels per acre)</td>
<td>24,134</td>
<td>172.32</td>
<td>39.59</td>
<td>21.00</td>
<td>287.80</td>
</tr>
<tr>
<td>Planting Density (1000s per acre)</td>
<td>24,134</td>
<td>27.92</td>
<td>1.61</td>
<td>18.25</td>
<td>33.41</td>
</tr>
<tr>
<td>Fertilizer (lb per acre)</td>
<td>24,134</td>
<td>132.67</td>
<td>44.42</td>
<td>0.50</td>
<td>236.25</td>
</tr>
<tr>
<td>Irrigation</td>
<td>24,134</td>
<td>0.10</td>
<td>0.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Insecticide</td>
<td>24,134</td>
<td>0.33</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7: Summary Statistics

### 4.2 Empirical Strategy

Our empirical analysis contains two parts. In the first part, we use the proposed nonparametric method and the traditional parametric method to test if the moments of the yields
from the selected GM corn varieties are significantly different from the conventional variety. This relies on the inference procedure on categorical variables described in Section 2.2.2. In the second part, we use the proposed inference procedure on continuous variable described in Section 2.2.1 to examine the effect of planting density on the moments of the yields. In both parts, we use Gaussian kernel function. The bandwidths for continuous variables are chosen by rule-of-thumb, and the bandwidths for categorical variables are chosen by cross validation. The number of bootstrap is 500.

In addition to the conditional moments, we also examine the cost of risk (as in Shi et al. (2013)), which is a summary measure of all order moments (first order to fourth order in our case). In the following, we define the cost of risk. Assume the utility function of the decision maker is constant relative risk aversion (CRRA) with parameter $r$, i.e.,

$$U(y) = y^{1-r}/(1-r).$$

The risk aversion parameter $r$ usually falls into the interval of $[1,5]$ (Gollier, 2004). We pick $r = 3$ for illustrative purpose. With this utility function, the cost of risk or risk premium can be written as,

$$R(x, z) = M_1(x, z) - U^{-1}[EU(y(x, z))].$$

By simple derivations, one can approximate the cost of risk by

$$R(x, z) \approx \frac{1}{\partial U/\partial y} \left[ - \frac{1}{2} \cdot \frac{\partial^2 U}{\partial y^2} M_2(x, z) - \frac{1}{6} \cdot \frac{\partial^3 U}{\partial y^3} M_3(x, z) - \frac{1}{24} \cdot \frac{\partial^4 U}{\partial y^4} M_4(x, z) \right],$$

where the derivatives (from first order to fourth order) of $U(\cdot)$ are evaluated at $M_1(x, z)$. Plugging in the specification of $U(\cdot)$ with $r = 3$, we have

$$R(x, z) \approx \frac{3}{2} \cdot \frac{M_2(x, z)}{M_1(x, z)} - 2 \cdot \frac{M_3(x, z)}{M_1^2(x, z)} + \frac{5}{2} \cdot \frac{M_4(x, z)}{M_1^3(x, z)}.$$

Note that the cost of risk is defined at a specific point of $(x, z)$. To have a comprehensive representation, in our empirical results, we integrate $x$ and $z$ out with their empirical distributions.

4.3 Results: Effects of GM Varieties

Table 8 reports the estimated differences between the moments of two specific GM varieties selected for this study and the corresponding moments of the conventional non-GM variety. The first two columns are the results for the ECB, while the last two columns are the results...
for ECB-GFT. The first and third columns are for parametric method, and the second and fourth columns are for nonparametric method. We consider three significance levels: 5%(*), 1%(**) and 0.1%(***).

We see that the parametric method works well only in estimating the first moment. It gives similar estimates (in magnitude and significance level) as the nonparametric method does. For higher order moments (higher than first moment), the parametric method gives no significant results, and hence we cannot draw any conclusions from the signs of the estimates. On the other hand, the nonparametric method gives statistically significant estimates. The GM varieties have higher first and third moments, and lower second and fourth moments than the conventional non-GM variety does. This is consistent with our theoretical expectation.

In addition, recall that in the Monte Carlo simulations (Section 3), we show that the nonparametric method has stronger test power than the parametric method. The empirical findings here are consistent with the simulation results.

With regards to the cost of risk, the parametric method gives a significant estimate only for ECB, while the nonparametric method gives significant estimates for both ECB and ECB-GFT. Researchers typically expect that when stacking genes, the cost of risk should decrease (i.e., that is, overall riskiness becomes smaller as one stacks more genes to protect against more pest risk). This suggests that farmers are willing to give up less and less bushels to replace a risky yield with the mean yield as one stacks more and more genetic traits that protects against different risks. The results from the nonparametric method support this hypothesis, while the results from the parametric method do not.

<table>
<thead>
<tr>
<th>Variety</th>
<th>ECB</th>
<th>ECB-GFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>Mean</td>
<td>8.53***</td>
<td>8.59***</td>
</tr>
<tr>
<td>Variance</td>
<td>-27.31</td>
<td>-61.73***</td>
</tr>
<tr>
<td>Skewness</td>
<td>834</td>
<td>938*</td>
</tr>
<tr>
<td>Kurtosis (×10³)</td>
<td>-214</td>
<td>-139***</td>
</tr>
<tr>
<td>Cost of Risk</td>
<td>-0.72*</td>
<td>-0.70***</td>
</tr>
</tbody>
</table>

Table 8: Estimated Differences in Moments from GM varieties vs. Conventional Variety

The original estimates (not difference) of conditional moments and costs of risk for all three varieties are reported in Table 10 in Appendix B.

### 4.4 Results: Effects of Planting Density

We focus on the ECB variety, and examine the effects of planting density on the integrated moments. We set the fertilizer at its mean value, and integrate out all other variables. Figure
I shows the relationship between each order moment and planting density. It displays a clear pattern that all order moments are positively correlated with planting density. We then run two tests to support such a pattern: (1) test if the conditional moment is increasing in planting density; (2) test if the conditional moment is decreasing in planting density. Table 9 shows the testing results. We see that for testing increasing, we cannot reject the null hypothesis in each case (first order to fourth order moments). For testing decreasing, we can reject the null hypothesis for conditional mean, variance and skewness, but cannot reject the null hypothesis for conditional kurtosis.

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing</td>
<td>Not reject</td>
<td>Not reject</td>
<td>Not reject</td>
<td>Not reject</td>
</tr>
<tr>
<td>Decreasing</td>
<td>Reject</td>
<td>Reject</td>
<td>Reject</td>
<td>Not reject</td>
</tr>
</tbody>
</table>

Table 9: Monotonicity Test Results on ECB Yield Data

Note that the cost of risk is decreasing in the first and third moments, and is increasing in the second and fourth moments. Figure 1 does not provide a clear pattern on the relationship between cost of risk and planting density. We then compute the cost of risk using the estimated conditional moments, and plot its relationship with planting density in Figure 2. We see that as the planting density increases from 24 to 28, the cost of risk increases. This is due to the rapid increases in second and fourth moments, as shown in Figure 1. When the planting density increases from 28 to 33, the cost of risk decreases. This is because the
second and fourth moments become flat. Figure 2 suggests that the optimal planting density regarding cost of risk is around 24.

![Cost of Risk versus Planting Density](image)

**Figure 2: Estimated Cost of Risk versus Planting Density**

## 5 Conclusion

This paper proposes a nonparametric approach for evaluating factors that influence production risk. Compared to the traditional parametric approach, the proposed nonparametric procedure does not suffer from the risk of mis-specification. Our approach can estimate any order conditional moment effect to satisfy various research needs when evaluating production risk. Unlike standard nonparametric methods, our method can smooth not only continuous regressors, but also categorical regressors, which is extremely useful in evaluating how a variety of variables affect production risk. In addition to the estimation, we also provide comprehensive statistical inference procedures, which work for both continuous and categorical variables.

Monte Carlo simulations were also conducted to assess the performance of the proposed nonparametric estimation and inference methods. On the estimation side, our nonparametric approach produces smaller IMSEs compared to the traditional parametric approach. On the inference side, we find that the traditional parametric approach suffers from test power loss.
when there is mis-specification, while our nonparametric approach is robust to any mis-specification error.

Using field trial data from Wisconsin, we also provide an empirical illustration of the proposed nonparametric procedure by exploring how GM corn varieties and other input variables influence corn production risk. Our results indicate that our proposed nonparametric method produces stronger and more theoretically consistent results (relative to the parametric approach). The GM variety is found to statistically affect more of the higher moments when using the nonparametric procedure as compared to the parametric one. This is likely due the stringent linearity assumption imposed in the parametric method.

Overall, our paper has illustrated how a newly developed nonparametric procedure can be used as an additional tool in the economic assessment of production risk. It extends the “toolbox” available for economists when assessing how categorical and continuous variables influence the higher moments of a stochastic production function. However, further refinements and extensions are still possible. First, extending the nonparametric procedure to be able to handle endogenous input variables is needed (i.e., instrumental variable type procedures). A number of input variables are typically endogenous and an approach to deal with this within the nonparametric framework will further enhance its usefulness. Second, a nonparametric procedure developed for panel data structures would also be an interesting extension. These topics are left for future research.
Appendix for
"Nonparametric Estimation and Inference of Production Risk"

A Proof of Proposition 1

In this proof, we establish the asymptotic properties of the estimators \( \hat{M}_j(x, z), j \geq 2 \). For notation simplicity, let

\[
K_i \equiv K_h(x_i - x), \quad K_{ji} \equiv K_h(x_j - x_i), \quad \Lambda_i \equiv \Lambda_\lambda(z_i, z), \quad \Lambda_{ji} \equiv \Lambda_\lambda(z_j, z_i),
\]

\[
\Delta_i \equiv g(x_i, z_i) - \hat{M}_1(x, z), \quad \eta_n \equiv ||h||^2 + ||\lambda|| + \frac{1}{\sqrt{nH}}.
\]

We write \( a_n \approx b_n \) if \( a_n = b_n(1 + o(1)) \), and \( a_n \overset{P}{\approx} b_n \) if \( a_n = b_n(1 + o_P(1)) \).

Consider the difference between the estimator \( \hat{M}_j(x, z) \) and the true value \( M_j(x, z) \).

\[
\hat{M}_j(x, z) - M_j(x, z) = \frac{1}{nH} \sum_{i=1}^{n} (y_i - \hat{M}_1(x_i, z_i)) j K_i \Lambda_i - M_j(x, z)
\]

\[
= \frac{1}{nH} \sum_{i=1}^{n} \left\{ (\Delta_i + u_i) j - M_j(x, z) \right\} K_i \Lambda_i
\]

\[
= \frac{1}{nH} \sum_{i=1}^{n} \left\{ \sum_{p=0}^{j} \binom{j}{p} \Delta_i^p u_i^{j-p} - M_j(x, z) \right\} K_i \Lambda_i
\]

\[
\overset{P}{\approx} \frac{1}{nH} \sum_{i=1}^{n} \left\{ \sum_{p=2}^{j} \binom{j}{p} \Delta_i^p u_i^{j-p} + j \Delta_i u_i^{j-1} + u_i^j - M_j(x, z) \right\} K_i \Lambda_i
\]

\[
= A_1 + j A_2 + A_3,
\]

where

\[
A_1 = \frac{1}{nH} \sum_{i=1}^{n} \left[ \sum_{p=2}^{j} \binom{j}{p} \Delta_i^p u_i^{j-p} \right] K_i \Lambda_i
\]

\[
A_2 = \frac{1}{nH} \sum_{i=1}^{n} \Delta_i u_i^{j-1} K_i \Lambda_i
\]

and

\[
A_3 = \frac{1}{nH} \sum_{i=1}^{n} \left[ u_i^j - M_j(x, z) \right] K_i \Lambda_i.
\]

In the next, we show that \( A_1 = o_p(\eta_n) \). By the uniform convergence established in Li
and Ouyang (2005), we have

$$
\sup_{(x,z) \in D_x \times D_z} |g(x, z) - \widehat{M}_1(x, z)| = O_p \left( \sum_{s=1}^{d_1} h_s^2 + \sum_{s=1}^{d_2} \lambda_s + \frac{\log(n)}{\sqrt{nH}} \right).
$$

(A.1)

As a result,

$$
\max_{1 \leq i \leq n} |\Delta_i^p| \leq \sup_{(x,z) \in D_x \times D_z} |g(x, z) - \widehat{M}_1(x, z)|^p = o_p(\eta_n), \quad p \geq 2.
$$

(A.2)

Meanwhile, we have that

$$
\frac{1}{nH} \sum_{i=1}^{n} \left[ \sum_{p=2}^{j} (j)_{p}^{-1} u_{i}^{j-p} \right] K_i \Lambda_i = O_p(1).
$$

(A.3)

Combining (A.2) and (A.3), we have $A_1 = o_p(\eta_n)$.

In the next, we show the asymptotic properties of $A_2$. When $j = 2$, using the uniform convergence in (A.1) and the fact that

$$
\frac{1}{nH} \sum_{i=1}^{n} u_i K_i \Lambda_i = O_p \left( \frac{1}{\sqrt{nH}} \right);
$$

it is easy to see that $A_2 = o_p(\eta_n)$. We discuss the case when $j = 2$ in detail at the end of the proof. Here, let $j \geq 3$.

$$
A_2 = \frac{1}{nH} \sum_{i=1}^{n} u_i^{j-1} (g(x_i, z_i) - \widehat{M}_1(x_i, z_i)) K_i \Lambda_i
$$

$$
\approx \frac{1}{f(x, z)} \cdot \frac{1}{n^2 H^2} \sum_{i=1}^{n} \frac{u_i^{j-1}}{f(x_i, z_i)} \sum_{j=1}^{n} \left[ g(x_i, z_i) - g(x_j, z_j) - u_j \right] K_j \Lambda_j K_i \Lambda_i
$$

$$
\approx \frac{1}{f(x, z)} \cdot \frac{1}{n^2 H^2} \sum_{i=1}^{n} \frac{M_{j-1}(x_i, z_i)}{f(x_i, z_i)} \sum_{j \neq i} \left[ g(x_i, z_i) - g(x_j, z_j) - u_j \right] K_j \Lambda_j K_i \Lambda_i
$$

$$
= A_{21} - A_{22},
$$

where

$$
A_{21} = \frac{1}{f(x, z)} \cdot \frac{1}{n^2 H^2} \sum_{i=1}^{n} \sum_{j \neq i} \frac{M_{j-1}(x_i, z_i)}{f(x_i, z_i)} \left[ g(x_i, z_i) - g(x_j, z_j) \right] K_j \Lambda_j K_i \Lambda_i
$$
and
\[ A_{22} = \frac{1}{f(x, z)} \cdot \frac{1}{n^2 H^2} \sum_{i=1}^{n} \sum_{j \neq i} M_{j-1}(x_i, z_i) u_j K_{ji} \Lambda_{ji} K_i \Lambda_i. \]

Note that both \( A_{21} \) and \( A_{22} \) can be written as second order U-statistics. By the U-statistics H-decomposition, it is easy to show that
\[ A_{21} = E(A_{21}) + o_p(\eta_n) = M_{j-1}(x, z) B_1(x, z) + o_p(\eta_n) \tag{A.4} \]
and
\[ A_{22} = \frac{1}{f(x, z)} \cdot \frac{1}{n H} \sum_{i=1}^{n} u_i M_{j-1}(x_i, z_i) K_i \Lambda_i + o_p(\eta_n), \tag{A.5} \]

where \( B_1(x, z) \) is defined in (2.2).

In the next, we show the asymptotic properties of \( A_3 \).

\[ A_3 \overset{p}{\approx} \frac{1}{n H} \sum_{i=1}^{n} [u_i M_j(x, z) - M_j(x, z)] K_i \Lambda_i \]
\[ = \frac{1}{n H} \sum_{i=1}^{n} [u_i M_j(x, z) + M_j(x, z) - M_j(x, z)] \frac{K_i \Lambda_i}{f(x, z)} \]
\[ = A_{31} + A_{32}, \]

where
\[ A_{31} = \frac{1}{f(x, z)} \cdot \frac{1}{n H} \sum_{i=1}^{n} [M_j(x, z) - M_j(x, z)] K_i \Lambda_i \tag{A.6} \]
and
\[ A_{32} = \frac{1}{f(x, z)} \cdot \frac{1}{n H} \sum_{i=1}^{n} [u_i M_j(x, z)] K_i \Lambda_i. \]

By standard arguments (e.g. Li and Racine, 2007), it is straightforward that
\[ A_{31} = B_j(x, z) + o_p(\eta_n) \tag{A.7} \]
where \( B_j(x, z) \) is defined in (2.2).
Combining \((A.4)-(A.7)\), we have

\[
\hat{M}_j(x, z) - M_j(x, z) = jA_{21} - jA_{22} + A_{31} + A_{32} + o_p(\eta_n)
\]

\[
= \frac{1}{f(x, z)} \cdot \frac{1}{nH} \sum_{i=1}^{n} \left[ -ju_iM_{j-1}(x_i, z_i) + u_i^j - M_j(x_i, z_i) \right] K_iA_i + jM_{j-1}(x, z)B_1(x, z) + B_j(x, z) + o_p(\eta_n)
\]

\[
= A_4 + \tilde{B}_j(x, z) + o_p(\eta_n)
\]

where

\[
A_4 = \frac{1}{f(x, z)} \cdot \frac{1}{nH} \sum_{i=1}^{n} \left[ -ju_iM_{j-1}(x_i, z_i) + u_i^j - M_j(x_i, z_i) \right] K_iA_i,
\]

and \(\tilde{B}_j(x, z)\) is defined in \((2.3)\). It is easy to see that \(E(A_4) = 0\) and

\[
\text{Var}(A_4) = \frac{\kappa_{d^1}}{nH f(x, z)} \left[ j^2M_2(x, z)M_{j-1}^2(x, z) + M_{2j}(x, z) - M_j^2(x, z) - 2jM_{j-1}(x, z)M_{j+1}(x, z) \right] + o_p\left(\frac{1}{nH}\right)
\]

\[
= \frac{1}{nH} \tilde{\Omega}_j(x, z) + o_p\left(\frac{1}{nH}\right),
\]

where \(\tilde{\Omega}_j(x, z)\) is defined in \((2.5)\). Then by Lyapunov’s Central Limit Theorem, we have that

\[
\sqrt{nH}(\hat{M}_j(x, z) - M_j(x, z) - \tilde{B}_j(x, z)) \stackrel{d}{\to} N(0, \tilde{\Omega}_j(x, z)), \quad j \geq 3.
\]

In the last, we examine the case when \(j = 2\). Since \(A_2 = o_p(\eta_n)\),

\[
\hat{M}_2(x, z) - M_2(x, z) = A_{31} + A_{32} + o_p(\eta_n)
\]

\[
= B_2(x, z) + A_{32} + o_p(\eta_n),
\]

where \(B_2(x, z)\) is defined in \((2.2)\). It is easy to see that \(E(A_{32}) = 0\) and

\[
\text{Var}(A_{32}) = \frac{1}{nH} \Omega_j(x, z) + o_p(\eta_n),
\]

where \(\Omega_j(x, z)\) is defined in \((2.4)\). Then by Lyapunov’s Central Limit Theorem, we have that

\[
\sqrt{nH}(\hat{M}_2(x, z) - M_2(x, z) - B_2(x, z)) \stackrel{d}{\to} N(0, \Omega_2(x, z)).
\]
### B Supplemental Tables

<table>
<thead>
<tr>
<th>Variety</th>
<th>Conventional</th>
<th>ECB</th>
<th>ECB-GFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>P</td>
<td>NP</td>
<td>P</td>
</tr>
<tr>
<td>Mean</td>
<td>189.13</td>
<td>189.82</td>
<td>197.67</td>
</tr>
<tr>
<td>Variance</td>
<td>613.37</td>
<td>249.09</td>
<td>586.06</td>
</tr>
<tr>
<td>Skewness</td>
<td>-2424</td>
<td>-1713</td>
<td>-1590</td>
</tr>
<tr>
<td>Kurtosis ($\times 10^3$)</td>
<td>1254</td>
<td>260</td>
<td>1040</td>
</tr>
<tr>
<td>Cost of Risk</td>
<td>6.24</td>
<td>2.37</td>
<td>5.52</td>
</tr>
</tbody>
</table>

Table 10: Estimated Moments of Conventional, ECB and ECB-GFT Varieties
References


