INFERENCE IN THE NONPARAMETRIC STOCHASTIC FRONTIER MODEL

Christopher F. Parmeter, Léopold Simar,
Ingrid Van Keilegom, Valentin Zelenyuk

LIDAM Discussion Paper ISBA
2021 / 29

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

INFERENCE IN THE NONPARAMETRIC STOCHASTIC FRONTIER MODEL

CHRISTOPHER F. PARMETER, LÉOPOLD SIMAR, INGRID VAN KEILEGOM, AND VALENTIN ZELENYUK

Abstract. This paper is the first in the literature to discuss in detail how to conduct various types of inference in the stochastic frontier model when it is estimated using nonparametric methods. We discuss a general and versatile inferential technique that allows for a range of practical hypotheses of interest to be tested. We also discuss several challenges that currently exist in this framework in an effort to alert researchers to potential pitfalls. Namely, it appears that when one wishes to estimate a stochastic frontier in a fully nonparametric framework, separability between inputs and determinants of inefficiency is an essential ingredient for the correct empirical size of a test. We showcase the performance of the test with a variety of Monte Carlo simulations.

Key Words: Stochastic Frontier Analysis, Efficiency, Productivity Analysis, Local-Polynomial Least-Squares.

JEL Classification: C1,C14, C13

Christopher F. Parmeter, Department of Economics, University of Miami, USA. e-mail: cparmeter@bus.miami.edu. Léopold Simar, Institut de statistique, biostatistique et sciences actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium. e-mail: leopold.simar@uclouvain.be. Ingrid Van Keilegom, ORSTAT, KU Leuven, Naamsestraat 69, B3000 Leuven, Belgium and Institut de statistique, biostatistique et sciences actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium. e-mail: ingrid.vankeilegom@kuleuven.be. Valentin Zelenyuk, School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Australia. e-mail: v.zelenyuk@uq.edu.au. Ingrid Van Keilegom acknowledges the financial support from the European Research Council (2016-2021, Horizon 2020/ERC grant agreement No. 694409). Valentin Zelenyuk acknowledges the financial support from Australian Research Council (FT170100401). Comments on earlier drafts from William Horrace, Subal Kumbhakar and Robin Sickles as well as feedback from Bao Hoang Nguyen, Zhichao Wang and Evelyn Smart helped improve the manuscript. The usual caveat applies.

1. Introduction

The stochastic frontier model (SFM) is one of the workhorse frameworks in applied efficiency analysis. However, it is routinely inveighed against for its imposition of rigid parametric assumptions which are typically difficult to justify. For example, specification of the frontier as Cobb-Douglas or translog is almost universally taken as given with little in the way of specification testing. Other parametric assumptions include normality of the error terms and a basic set of distributions for the one-sided inefficiency component (this is usually done for analytic convenience rather than an unwavering belief in the correct parametric shape of a given distributional family). Beyond this trinity of parametric assumptions, researchers have also injected strict functional forms into the behavior of how determinants of inefficiency influence the shape and location of the inefficiency distribution. Needless to say, the use of parametric assumptions at all modeling stages for the SFM are legion. Robust and consistent tests of specification and/or significance have yet to be roundly advocated for in the literature. We take a step in this direction here.

With tongue-in-cheek, nonparametric specification testing of the SFM represents one of the last frontiers in this literature. Recent advances in nonparametric estimation (Kumbhakar, Park, Simar & Tsionas 2007, Kneip, Simar & Van Keilegom 2015), panel data modeling (Greene 2005, Wang & Ho 2010, Chen, Schmidt & Wang 2014), handling endogeneity (Amsler, Prokhorov & Schmidt 2016, Amsler, Prokhorov & Schmidt 2017, Kutlu, Tran & Tsionas 2019), the use of quantiles (Behr 2010), spatial spillovers (Glass, Kenjegalieva & Sickles 2016, Orea & Álvarez 2019) and treatment effects (Chen, Hsu & Wang 2020) have all made their way to, or are progressing towards, maturity, as is more recent work on general issues specific to the SFM (Belotti & Ilardi 2018, Horrace & Wright 2020).

However, at present, a versatile testing framework that is nonparametric in nature has remained elusive. Indeed, virtually all studies that have proposed semi- or nonparametric approaches for the SFM have almost exclusively focused on aspects of estimation, leaving questions of inference for future research or discussing them in mere passing at best. As a case in point, consider the recent review of semi- and nonparametric approaches for the SFM in Parmeter & Zelenyuk (2019). While Parmeter & Zelenyuk (2019) dedicated a separate section to discuss various possibilities on how inference for such estimation approaches could potentially be done, their Monte Carlo focused only on estimation. This is largely because the research on the inference in such models, even with one particular approach, turned into a long-lasting study by itself, a summary of which is this very paper.

Jumping ahead, the main conclusion of this study is that conducting nonparametric inference, often portrayed as the main advantage of the SFM over data envelopment analysis (DEA), turns out not to be as simple or straightforward as perhaps has been suggested in the literature. Specifically, for nonparametric inference, a carefully-designed bootstrap is commonly required as asymptotic methods are known to perform poorly in finite samples[1] and are needed for appropriate construction of a test for correct specification in general, and especially for the SFM with convoluted error terms and unknown functions for both the frontier and the mean of inefficiency. More importantly, our simulations here reveal that the so-called separability assumption arises as a crucial requirement for our nonparametric tests to perform optimally, just as DEA requires it for the two-stage truncated regression approach of Simar & Wilson (2007). Thus, while solid nonparametric inference can be performed in the SFM, our work here suggests that many preconceived notions of dominance with respect to inference appear overstated at least in the nonparametric setting.

While there are many testing alternatives, we will focus on an omnibus test for many of the most crucial and important hypotheses that surround production and cost frontiers.[2] This test is based on the seminal work of Delgado & González Manteiga (2001) and can also be viewed as further generalization of Kim & Schmidt (2008) who considered inference for parametric SFMs. However, direct application of their test is not available in the nonparametric SFM. This stems from the fact that in the fully nonparametric setting, the location of the frontier is corrupted by expected inefficiency, which must first be removed prior to inference being conducted (either on the frontier directly, or the conditional mean of inefficiency).

For testing in a nonparametric regression context, various approaches have been suggested in the general statistical/econometric literature: e.g., Ullah (1985), Härdle & Mammen (1993), Fan & Li (1996), Zheng (1996), Li & Wang (1998), Delgado & González Manteiga (2001) and Fan, Zhang & Zhang (2001), to mention just a few. Our focus will be on adaptation of the tests from Delgado & González Manteiga (2001), hereafter DGM. While other approaches could be deployed, for example, conditional moment tests such as Zheng (1996), the main discussion and implementation we discuss here is likely to carry over analogously in these settings.

A battery of parametric specification tests are proposed as well as two of the most prominent applied tests are put under computational scrutiny: tests of significance and a test of

[1]See Härdle & Mammen (1993) for one of the first discussions on this as well as more recent textbook discussions in Pagan & Ullah (1999) and Henderson & Parmeter (2015).
[2]And of course input and output distance functions, profit frontiers, revenue frontiers, etc.

correct parametric specification. These simulations reveal two key points. First, when we have the so-called separability property (Simar & Wilson 2007), the DGM test works quite well (see Section 4.2). Second, when we do not have separability, the test fails; failure here refers to the empirical size of the test deviating substantially from the nominal size (see Section 4.1).

We believe that the reason inference in nonparametric SFMs is likely to remain difficult stems from the fact that when the frontier and inefficiency contain the same variables (separability fails), this makes it much harder on the nonparametric test. This is due to the additional noise that is introduced via estimation of both the frontier and the conditional mean of inefficiency given that the covariates affect both. We also note that this issue is not specific to the exact form of the DGM test statistic, but is more generally related to how the frontier and conditional mean of inefficiency are so difficult to untangle without other specific assumptions (separability between traditional inputs and environmental variables for example).

To further examine this we have also explored the question of how would a parametric estimator perform in this case instead of a nonparametric estimator and we have reached a somewhat unexpected conclusion. Even a fully parametric stochastic frontier model, estimated via nonlinear least squares (NLS) under correct specification, shows a similarly poor performance in terms of the size of the test (see Figure 3), and this is when the starting values in the optimization were set to the true values of the parameters. Hence, hoping that the preferred nonparametric estimator (and likely other nonparametric approaches) would perform well in terms of size, even when a correctly specified parametric approach cannot do so, appears to be unrealistic. In turn, this justifies the approach of Kim & Schmidt (2008) who only considered the case with separability in their Monte Carlo scenarios and theoretical derivations.

Given this, our work here can be viewed as the nonparametric generalization of Kim & Schmidt (2008), which is currently state of the art for inference in the parametric SFM. Our goal is to both generalize their testing environment to the nonparametric setting and enhance the suite of hypotheses that can be examined. A key aspect of Kim & Schmidt (2008) is that the array of tests that they propose hinge critically on the production frontier being correctly specified and are focused exclusively on testing significance of a given set of characteristics on efficiency. In our testing framework, given that the conditional mean is estimated nonparametrically, no such reliance is present. Moreover, the framework of

DGM allows for a broader range of hypotheses to be tested. Our work is also analogous or 'parallel' to the work of Simar & Wilson (2007) in the DEA context, reaching a similar general conclusion that nonparametric inference is quite challenging, yet possible under the separability assumption along with an appropriate bootstrap.

The remainder of the paper is structured as follows. Section 2 briefly reviews the most recent techniques for nonparametric estimation of both the production frontier and the conditional mean of inefficiency. Section 3 discusses the general framework of the test of Delgado & González Manteiga (2001) as well as adaptations to it that specifically deploy hypotheses of interest for SFM. Section 4 presents a detailed set of simulations surrounding the performance of the DGM test for both statistical significance and correct parametric specification. Lastly, Section 5 discusses empirical recommendations based on the theoretical and simulated insights generated here.

2. Nonparametric Estimation of the Stochastic Frontier Model

Consider a set of i.i.d. random variables (X_i, Z_i, Y_i), for $i = 1, \ldots, n$, where $X_i \in \mathbb{R}^p$ is a vector of traditional inputs used to produce the vector of outputs $Y_i \in \mathbb{R}$, while vector $Z_i \in \mathbb{R}^d$ represents a separate set of variables that may influence production, sometimes referred to as 'environmental variables'.[3]

We assume that the data generating process is characterized by the joint pdf of (X, Z, Y) that can be decomposed into a joint marginal for (X, Z) and a conditional pdf for Y given (X, Z), so that the conditional of Y given $X = x$ and $Z = z$ is characterized through

$$Y = m(x, z) - U + V, \tag{1}$$

where $m(x, z)$ is the production frontier, $V|X = x, Z = z \sim D(0, var_V(x, z))$ where $D(0, \cdot)$ is a real random variable with mean zero and some positive and finite variance $var_V(\cdot, \cdot)$ and $U|X = x, Z = z \sim D^+(\mu_U(x, z), var_U(x, z))$ where $D^+(\cdot, \cdot)$ is a positive random variable with mean $\mu_U(\cdot, \cdot)$ and variance $var_U(\cdot, \cdot)$. We also suppose that, conditionally on (X, Z), U and V are independent random variables, where V has a symmetric distribution around zero, and U is a non-negative random variable from a right-skewed distribution. Note that some

[3]The distinction between X_i and Z_i in some contexts might be unimportant from a statistical perspective, yet it might be imperative from an economic perspective and therefore we keep them separate. E.g., inputs are usually considered as factors under the control of the firm for producing outputs, while the environmental variables are often viewed as those that may influence production but are not under the control of the firm (e.g., geography, regulatory conditions, climate, etc.).

existing approaches restrict the channels through which Z can influence output, typically by imposing some structure a priori on the relationship, for example, the so-called 'separability condition' whereby $m(x,z) = m(x)$ and $\mu_U(x,z) = \mu_U(z)$ and $var_U(x,z) = var_U(z)$ (see Simar & Wilson 2007, for related discussion).

Unlike parametric approaches, we allow the production frontier $m(x,z)$ to be completely *unknown*, where the goal of the researcher is to both estimate and perform inference about the production technology (scale elasticities, marginal productivity of inputs, etc.) and inefficiency. In particular a researcher may be specifically interested in how inefficiency is related to (x,z), by utilizing the sample of i.i.d. triples $\mathcal{S}_n = \{(X_i, Z_i, Y_i) \,|\, i = 1, \ldots, n\}$. While several approaches exist to estimate the SFM in a nonparametric fashion (Fan, Li & Weersink 1996, Kumbhakar et al. 2007, Martins-Filho & Yao 2015, Park, Simar & Zelenyuk 2015)[4] here we advocate the use of the approach proposed in Simar, Van Keilegom & Zelenyuk (2017), hereafter SVKZ, which is a nonparametric generalization of Olson, Schmidt & Waldman (1980). In SVKZ, as with the vast majority of papers in this area, interest has mainly hinged on the development of estimators of the SFM and their asymptotic properties. Our focus here will be exclusively on inference predicated on nonparametric estimation of this model. The findings detailed here are likely to have practical implications for many of the previous approaches, a focused investigation which we leave for future research.

Our focus on the nonparametric equivalent of the initial corrected ordinary least-squares approach of Olson et al. (1980) stems from the fact that the existing approaches all have what we consider to be practical limitations that restrict the ability for general inference in practice. Both Fan et al.'s (1996) and Martins-Filho & Yao's (2015) approaches do not allow one to model the impact of (X,Z) on inefficiency while Kumbhakar et al. (2007) and Park et al. (2015) rely on local maximum likelihood, which can be quite computationally burdensome. On the contrary, SVKZ's corrected local least-squares approach can be rapidly deployed and is computationally simple. This makes it a convenient candidate to discuss general inference.

To describe how to conduct inference in SVKZ's framework, we first summarize the key points of their approach which consists of three stages. The first stage is dedicated to estimation of the conditional mean rather than the frontier. This is due to the fact that $\mathbb{E}(U|x,z) \neq 0$ and hence is conflated with the true frontier when estimated directly. Defining

[4]See Parmeter & Zelenyuk (2019) for a recent review.

$\varepsilon = V - U + \mu_U(x,z)$, and $r_1(x,z) = m(x,z) - \mu_U(x,z)$, one can rewrite Equation (1) as

$$\begin{aligned} Y =& m(x,z) - \mu_U(x,z) + V - U + \mu_U(x,z) \\ =& r_1(x,z) + \varepsilon. \end{aligned} \tag{2}$$

Note that $\mathbb{E}(\varepsilon|x,z) = 0$ and $\mathbb{V}(\varepsilon|x,z) = var_U(x,z) + var_V(x,z) \in (0,\infty)$, where $var_U(x,z)$ ($var_V(x,z)$) is a function of (x,z) which represents the conditional variance of $U|x,z$ ($V|x,z$). As $r_1(x,z) = \mathbb{E}(Y|x,z)$, Equation (2) can be estimated via standard nonparametric regression methods from $\mathcal{S}_n$. A promising candidate is local-linear least-squares, which has good asymptotic properties and is relatively easy to implement; the gradients of the conditional mean are also provided directly with this estimator.

Note that $\widehat{r}_1(x,z)$ is an estimate of $r_1(x,z) = m(x,z) - \mu_U(x,z)$, rather than of the production frontier $m(x,z)$ and because $\mu_U(x,z) \geq 0$, we have $r_1(x,z) \leq m(x,z)$, $\forall (x,z) \in \mathbb{R}^{p+d}$. Thus, $r_1(x,z) = m(x,z)$, $\forall (x,z) \in \mathbb{R}^{p+d}$, if and only if $\mu_U(x,z) = 0$. And, if there is inefficiency, then $\widehat{r}_1(x,z)$ would be a downward-biased estimator of $m(x,z)$ and the bias is exactly $\mu_U(x,z)$, and so it is potentially varying with (x,z). In the particular case when $\mu_U(x,z)$ is a constant (i.e., $\mathbb{E}(U|x,z) = \mathbb{E}(U)$), much of the characteristics of $r_1(x,z)$ are directly transferred to $m(x,z)$, except its location.

So, from a sample of i.i.d. data $\{(X_i, Z_i, Y_i) \ : \ i = 1, \ldots, n\}$ one can obtain the local-linear least-squares (LLLS) estimate of $r_1(x,z)$ by solving, for any point of interest (x,z)

$$(\hat{\alpha}_{x,z}, \hat{\beta}_{x,z}) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \left[Y_i - (\alpha + \beta'((X_i, Z_i) - (x,z)))\right]^2 K\left(\frac{X_i - x}{h_{1x}}, \frac{Z_i - z}{h_{1z}}\right), \tag{3}$$

where, with a slight abuse of notation, $K\left(\frac{X_i - x}{h_{1x}}, \frac{Z_i - z}{h_{1z}}\right)$ stands for a product kernel for the $p+d$ components of (X,Z) with $h_1 = (h_{1x}, h_{1z})$ denoting the vector of $p+d$ bandwidths (where the 1 in the subscript is used to signify that it is a vector of bandwidths for estimating r_1). Solving (3), yields

$$\begin{aligned} \widehat{r}_1(x,z) =& \hat{\alpha}_{x,z}, \\ \widehat{\nabla} r_1(x,z) =& \hat{\beta}_{x,z}, \end{aligned}$$

where the second equation provides an estimate of the gradient of $r_1(x,z)$ at (x,z). The bandwidths can be selected by the leave-one-out least-squares cross-validation (LSCV).

Note that the LLLS approach does not require nonlinear optimization (except for LSCV), since the solution to (3) can be obtained in closed form via simple linear algebra (see Li

& Racine 2007). Under fairly mild regularity conditions and an appropriate choice of the bandwidths, these estimators have desirable theoretical properties (consistency, asymptotic normality, etc., see e.g. Fan & Gijbels 1996).

The second stage of the approach of SVKZ is dedicated to the estimation of the second and third conditional moments of ε. This can be also done without particular assumptions for the local distributions of U and of V. Given the rearrangement of the frontier model in Equation (2), $\mathbb{E}(\varepsilon|x,z)=0$. Further, due to the assumed symmetry for V, we have

$$\mathbb{E}(\varepsilon^2|x,z) := r_2(x,z) = var_U(x,z) + var_V(x,z) > 0,$$
$$\mathbb{E}(\varepsilon^3|x,z) := r_3(x,z) = -\mathbb{E}\left[\big(U-\mu_U(x,z)\big)^3|x,z\right].$$

SVKZ proposes to estimate these moments nonparametrically, utilizing the residuals from Equation (2):

$$\widehat{\varepsilon}_i = Y_i - \widehat{r}_1(X_i,Z_i),\ i=1,\ldots,n,$$

where $\widehat{r}_1(X_i,Z_i)$ is the selected nonparametric estimator of the conditional mean in the first stage (e.g., the local-linear least-squares estimator). In this case, nonparametric regression of $\widehat{\varepsilon}^2$ on (X,Z) and $\widehat{\varepsilon}^3$ on (X,Z) will produce consistent estimators for $\mathbb{E}(\varepsilon^2|x,z)$ and $\mathbb{E}(\varepsilon^3|x,z)$, respectively.

The regression functions $r_2(x,z)$ and $r_3(x,z)$ can be consistently estimated from $\{(\widehat{\varepsilon}_i^2, X_i, Z_i)|i=1,\ldots,n\}$ and $\{(\widehat{\varepsilon}_i^3, X_i, Z_i)|i=1,\ldots,n\}$, via

$$\widehat{r}_2(x,z) = \sum_{i=1}^{n} A_{i,h_2}(x,z)\,\widehat{\varepsilon}_i^2 = \sum_{i=1}^{n} A_{i,h_2}(x,z)\big(Y_i - \widehat{r}_1(X_i,Z_i)\big)^2$$

and

$$\widehat{r}_3(x,z) = \sum_{i=1}^{n} A_{i,h_3}(x,z)\,\widehat{\varepsilon}_i^3 = \sum_{i=1}^{n} A_{i,h_3}(x,z)\big(Y_i - \widehat{r}_1(X_i,Z_i)\big)^3,$$

where $A_{i,h_j}(x,z)$ is short-hand notation for the corresponding elements of the matrix stemming from local-linear estimation of the model as detailed in Equation (3).

The third stage entails making a distributional assumption on U so that $\mu_U(x,z)$ can be estimated (no distributional assumption is required for V). SVKZ considered the popular choice of doing this through a Half Normal assumption for U. Given that the single parameter of U potentially depends on x and z, this makes the parametric assumption "local" (Simar

& Wilson 2021). Here we have

$$U|x,z \sim N_+\left(0, \sigma_U^2(x,z)\right), \tag{4}$$

for which we have

$$\mu_U(x,z) = \sqrt{2/\pi}\sigma_U(x,z)$$

and an estimate of σ_U is available from $\widehat{r}_3(x,z)$ as (see Olson et al. 1980)

$$\widehat{\sigma}_U(x,z) = \max\left\{0, \left[\sqrt{\frac{\pi}{2}}\left(\frac{\pi}{\pi-4}\right)\widehat{r}_3(x,z)\right]^{1/3}\right\}.$$

An alternative to this censoring would be to deploy the method of Hafner, Manner & Simar (2018). We leave a more robust discussion of this method for future research.

3. Inference

One of the main interests of the model and the method inspired by SVKZ is to allow a flexible modelization and an easy way to estimate the components of the models. However, practitioners like to test whether some elements of x or z are really significant in the process of producing inefficiency, or if the inputs x are all significant in describing the production frontier, or if the model for the frontier, or the model for the mean inefficiency could be simplified in some suitable parametric models. There are many possible settings a practitioner might want to test in our SFM framework.

We will focus our presentation on two general situations: (i) one could test whether some particular inputs or environmental variables influence the conditional mean of inefficiency (significance test); (ii) one could test whether some particular (parametric) functional form $m_0(\cdot)$ is appropriate to fit the data (specification test).[5]

For the first family of tests we will test the null hypothesis

$$H_0 : \mathbb{E}(U \mid X, Z) = \mathbb{E}(U \mid X_1, Z_1), \quad \text{a.s.}, \tag{5}$$

where $W_1 = (X_1, Z_1) \in \mathbb{R}^{p_1+d_1}$ is a subset of the full vector $W = (X, Z) \in \mathbb{R}^{p+d}$. Note that, as particular cases, we may have several combinations with $p_1 = 0$ or $p_1 = p$ and $d_1 = 0$ or

[5]Note that if we want to test whether all the inputs are relevant in the production frontier, we will show that it is easy to modify the significance test to the frontier model. Similarly, we may adapt the specification test for the conditional mean of the inefficiency, see Section 3.3 below.

$d_1 = d$. For the specification test we will test

$$H_0 : m(W) = m_0(W), \quad \text{a.s.},$$

where $m_0(\cdot)$ is some parametric model characterized by some finite-dimensional vector of unknown parameters, say θ.

We will adapt the approach of DGM to our nonparametric SFM. The DGM test is easy to implement and has a great flexibility allowing testing of various restrictions in the nonparametric regression framework. We note that an advantage of the DGM tests is that, in regular situations, they only require the estimation of the regression function under the null. We will see that, unfortunately, we lose a part of this peculiarity in our framework of SFM. The essence of the DGM approach, in standard regression models, where a dependent variable of interest, say G, is regressed over W, is to build a random variable $T(W)$, based on the difference between the weighted integrated regression function of G under the unrestricted model and the regression of G under the null. The selection of G and the definition of the variable T depend on the particular setup of the test, as will be illustrated in the next subsections. DGM show that, for an appropriate choice of the weighting function, the null hypothesis can be equivalently written as

$$H_0 : T(W) = 0, \quad \text{a.s.}.$$

Finally they propose two test statistics based on a suitable functional of an estimate $T_n(\cdot)$ of $T(\cdot)$. The two test statistics are the Cramér-von Mises (CM) statistic and the Kolmogorov-Smirnov statistic. The p-values, in both cases, are obtained using a bootstrap method. DGM's simulations reveal that both test statistics perform roughly equivalently. For the remainder of this paper we will deploy the CM version of the DGM test which is defined as

$$C_n = \sum_{i=1}^{n} T_n^2(W_i),$$

where the explicit formulation of $T_n(W)$ depends on the test and will be given below.

We will see below that the main difficulty in our setup of SFM, is that the dependent variable G needed to construct the test statistics is not observed and has to be estimated from the data. We will describe how this "contamination" can perturb the original variable $T(W)$ and its estimate.

3.1. **Significance test for inefficiency.** We remind the reader that we want to test

$$H_0 : \mathbb{E}(U \mid W) = \mathbb{E}(U \mid W_1), \quad \text{a.s.},$$

where $W_1 = (X_1, Z_1) \in \mathbb{R}^{p_1+d_1}$ is a subset of the full vector $W = (X, Z) \in \mathbb{R}^{p+d}$. The dependent variable of interest G is here U but the test described in DGM cannot be implemented, as such, because we do not observe the variable U. Otherwise the procedure would have been straightforward.

We define the random variable $T(W)$ as follows: for all $w = (x, z)$ define

$$T(w) = \mathbb{E}\left[f(X_1, Z_1)\big(U - \mathbb{E}(U|X_1, Z_1)\big)\mathbb{1}(W \leq w)\right].$$

Hence, H_0 is equivalent to $H_0 : T(W) = 0$, a.s., and an estimator of $T(w)$ is given by

$$T_n(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(X_{1,i}, Z_{1,i})\big(U_i - \widehat{\mathbb{E}}(U|X_{1,i}, Z_{1,i})\big)\mathbb{1}(W_i \leq w),$$

where $\widehat{\mathbb{E}}(U|X_{1,i}, Z_{1,i})$ is an appropriate nonparametric estimator of $\mathbb{E}(U|X_{1,i}, Z_{1,i})$ based on the sample $\{(U_i, X_{1,i}, Z_{1,i})\}_{i=1}^n$, and $\widehat{f}$ is an appropriate estimator of the density f of (X_1, Z_1).

Since U is not observed, this cannot be used, but if we assume that the density of U given $X = x, Z = z$ belongs to the one-parameter scale family, as in (4), we have by the symmetry of V that $\mu_U(x, z) = C\left(r_3(x, z)\right)^{1/3}$, where $r_3(x, z)$ is defined by

$$\varepsilon^3 = r_3(x, z) + \zeta, \text{ where } \mathbb{E}(\zeta|x, z) = 0.$$

Since for all (x, z), $\mathbb{E}\big(\varepsilon^3|x, z\big) = r_3(x, z)$, the null hypothesis (5) can equivalently be written as

$$H_0 : \mathbb{E}\big(\varepsilon^3|X, Z\big) = \mathbb{E}\big(\varepsilon^3|X_1, Z_1\big), \quad \text{a.s.}, \tag{6}$$

i.e., we are testing the same restriction, but in another regression, so that the analog of the previous test can be used replacing U by ε^3. Note that here we do not need the value of the constant C, so we do not need to specify which member of the one parameter scale family is chosen for $U|x, z$. However, ε_i^3 are not observed either, but they can be estimated by

$$\widehat{\varepsilon}_i^3 = \left[Y_i - \widehat{r}_1(X_i, Z_i)\right]^3,$$

where $\widehat{r}_1$ is an appropriate estimator of r_1. We define, as above for U, the random variable $T(w)$ by

$$T(w) = \mathbb{E}\left[f(X_1, Z_1)\big(\varepsilon^3 - \mathbb{E}(\varepsilon^3|X_1, Z_1)\big)\mathbb{1}(W \leq w)\right], \tag{7}$$

for all $w = (x, z)$ and again H_0 is equivalent to $H_0 : T(W) = 0$, a.s. Then, a test statistic could be derived by using an estimator of $T(w)$:

$$T_n(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(X_{1,i}, Z_{1,i})\big(\widehat{\varepsilon}_i^3 - \widehat{\mathbb{E}}(\widehat{\varepsilon}^3|X_{1,i}, Z_{1,i})\big)\mathbb{1}(W_i \leq w), \tag{8}$$

where $\widehat{\mathbb{E}}(\widehat{\varepsilon}^3|X_{1,i}, Z_{1,i})$ is an appropriate non-parametric estimator of $\mathbb{E}(\widehat{\varepsilon}^3|X_{1,i}, Z_{1,i})$, based on the sample $\{(\widehat{\varepsilon}_i^3, X_{1,i}, Z_{1,i})\}_{i=1}^n$.

Finally, the Cramér-von Mises test statistic is given by

$$C_n = \sum_{i=1}^{n} T_n(W_i)^2,$$

whose distribution is approximated by the wild bootstrap described in DGM, leading to the p-value, $\#\{C_n^* \geq C_n\}/B$, where C_n^* are the bootstrap realizations of C_n and B is the number of bootstrap repetitions.

The bootstrap algorithm follows as:

(1) Compute the residuals $\eta_i = \widehat{\varepsilon}_i^3 - \widehat{\mathbb{E}}(\widehat{\varepsilon}^3|X_{1,i}, Z_{1,i})$ and build a wild-bootstrap version, denoted by η_i^*.[6] This leads to the following bootstrap version of $\widehat{\varepsilon}_i^3$:

$$\widehat{\varepsilon}_i^{3,*} = \widehat{\mathbb{E}}(\widehat{\varepsilon}^3|X_{1,i}, Z_{1,i}) + \eta_i^*, \text{ for } i = 1, \ldots, n.$$

(2) Compute the bootstrap version of $T_n(w)$:

$$T_n^*(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(X_{1,i}, Z_{1,i})\big(\widehat{\varepsilon}_i^{3,*} - \widehat{\mathbb{E}}^*(\widehat{\varepsilon}^{3,*}|X_{1,i}, Z_{1,i})\big)\mathbb{1}(W_i \leq w),$$

where $\widehat{\mathbb{E}}^*(\widehat{\varepsilon}^{3,*}|X_{1,i}, Z_{1,i})$ is the nonparametric estimate of $\mathbb{E}(\widehat{\varepsilon}^{3,*}|X_{1,i}, Z_{1,i})$ in the bootstrap world, i.e., computed from the sample $\left\{(\widehat{\varepsilon}_i^{3,*}, X_{1,i}, Z_{1,i})\right\}_{i=1}^n$.

(3) Finally, the bootstrap version of the test statistic C_n is constructed as $C_{n,b}^* = \sum_{i=1}^n T_n^*(W_i)^2$.

(4) Repeat steps (2) and (3) B times.

[6] We can e.g., use the Rademacher procedure: we center the residuals to provide $\eta_i^c = \eta_i - n^{-1}\sum_{j=1}^n \eta_j$ and define $\eta_i^* = \eta_i^c \mathbb{B}_i - \eta_i^c(1 - \mathbb{B}_i)$ where $\mathbb{B}_i$ is a random draw from a Bernoulli variable $\mathbb{B}_i \in \{0, 1\}$ with probability of success $1/2$.

(5) Use the B bootstrap estimates from Step (4), to compute the bootstrap-based p-value, given by

$$\text{p-value} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left(C_{n,b}^{*} \geq C_n\right).$$

The consistency of the test will be based on the properties of the difference between $T(w)$ in (7) and our estimator $T_n(w)$ given in (8). The theory of the DGM test still applies under suitable conditions on the kernel smoothing function and bandwidth. The Monte-Carlo experiments in Section 4 confirm that the test behaves well, both for the achieved size and for the power when separability holds.

3.2. **Specification test.** We remind the reader that we want to test

$$H_0 : m(W) = m_0(W), \quad \text{a.s.},$$

where $m_0(\cdot)$ is some parametric model characterized by some finite-dimensional vector of unknown parameters, say θ. Another rewriting of (2) leads to define $\check{Y}$ by

$$\check{Y} = Y + \mu_U(x,z) = m(x,z) + \varepsilon.$$

Defined like this, the values of $\check{Y}$ are dispersed around the frontier level $m(x,z)$, according to the random deviations $\varepsilon = V - U + \mu_U(x,z)$ depending on (x,z) but with $\mathbb{E}(\varepsilon|x,z) = 0$.

So the DGM approach for testing H_0 would be easy if the dependent variable of interest $\check{Y}$ were observable. But the values of $\check{Y}$ cannot be observed, so we will replace them by the following estimate:

$$\begin{aligned} \widehat{\check{Y}} &= Y + \widehat{\mu}_U(x,z) \\ &= \check{Y} + \left(\widehat{\mu}_U(x,z) - \mu_U(x,z)\right) \\ &= m(x,z) + \left(\widehat{\mu}_U(x,z) - \mu_U(x,z)\right) + \varepsilon. \end{aligned} \tag{9}$$

Under the additional assumption of a one parameter scale family for $(U|x,z)$, we may obtain an estimator of $\mu_U(x,z)$ by the regression of $\widehat{\varepsilon}_i^3$ on (X_i, Z_i) leading to $\widehat{r}_3(x,z)$. Typically we have

$$\widehat{\mu}_U(x,z) = C\left(\widehat{r}_3(x,z)\right)^{1/3},$$

where C is some known constant depending on the choice of the family for $(U|x,z)$. Note that in this case we have to know the constant C.

Hence the DGM procedure to test H_0 is straightforward. In summary, the random variable of interest $T(w)$ is defined by:

$$T(w) = \mathbb{E}\left[f(W)\big(\check{Y} - m_0(W)\big)\mathbb{1}(W \leq w)\right],$$

for all $w = (x, z)$, where f is now the density of $W = (X, Z)$. An estimate of $T(w)$ is provided by

$$T_n(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(W_i)\big(\widehat{\check{Y}}_i - \widehat{m}_0(W_i)\big)\mathbb{1}(W_i \leq w), \tag{10}$$

where $\widehat{m}_0$ is a parametric estimator of m_0 based on the sample $\{(\widehat{\check{Y}}_i, X_i, Z_i)\}_{i=1}^n$.

The bootstrap analog of $\widetilde{T}_n(w)$ is similar as in the preceding case. Define

$$\widetilde{T}_n^*(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(W_i)\big(\widehat{\check{Y}}_i^* - \widehat{m}_0^*(W_i)\big)\mathbb{1}(W_i \leq w),$$

where $\widehat{\check{Y}}_i^* = \widehat{m}_0(W_i) + \widehat{\eta}_i^*$, with $\widehat{\eta}_i^*$ being a wild-bootstrap version of $\widehat{\eta}_i = \widehat{\check{Y}}_i - \widehat{m}_0(W_i))$ and $\widehat{m}_0^*(W_i)$ is the parametric estimator of m_0 from the bootstrap sample $\{(\widehat{\check{Y}}_i^*, X_i, Z_i)\}_{i=1}^n$.

In practice we will estimate the quantity f as follows:

$$\widehat{f}(w) = (n|h|)^{-1}\sum_{i=1}^{n} K((w_i - w)/h),$$

where $|h| = h_1 \cdots h_{p+d}$ and $K((w_i - w)/h)$ is our shorthand for the product kernel. We know that the additional noise introduced in the DGM test is coming from the term $\widehat{\mu}_U(x, z) - \mu_U(x, z)$ in (9). This term is again of order $O_p((nh^{p+d})^{-1/2})$ and we conjecture that the theory of the DGM test remains valid. Here too, the Monte-Carlo experiments in Section 4 confirm that the test behaves well, both for the achieved size and for the power.

3.3. **Testing other restrictions in the SFM.** Other restrictions could be tested by adapting the procedure described in the preceding sections. For instance we might test some parametric specification for the conditional mean of the inefficiency:

$$H_0 : \mathbb{E}(U \mid W) = g_0(W), \quad \text{a.s.}, \tag{11}$$

where $g_0(\cdot)$ is some parametric model. For the chosen one-parameter scale family for the conditional density of $U|x, z$, we have for the unrestricted case, $\mu_U(x, z) = C(r_3(x, z))^{1/3}$, where C is a known constant. As shown above in (6), we have the following analog of (11)

in terms of ε^3:

$$H_0 : \mathbb{E}(\varepsilon^3 \mid X, Z) = g_{\varepsilon,0}(X, Z), \quad \text{a.s.},$$

where $g_{\varepsilon,0}(x, z) = (g_0(x, z)/C)^3$ for all (x, z).

In the DGM approach the dependent variable of interest is ε^3 and the quantity of interest is

$$T(w) = \mathbb{E}\left[f(W)\big(\varepsilon^3 - g_{\varepsilon,0}(W)\big)\mathbb{1}(W \leq w)\right],$$

which can be estimated by

$$T_n(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(W_i)\big(\widehat{\varepsilon}_i^3 - \widehat{g}_0(W_i)\big)\mathbb{1}(W_i \leq w),$$

where $\widehat{g}_{\varepsilon,0}$ is a parametric estimator of $g_{\varepsilon,0}$ from the sample $\{(\widehat{\varepsilon}_i^3, X_i, Z_i)\}_{i=1}^n$. Then the test and the bootstrap can be implemented along the same lines as in the preceding subsections.

Another restriction that may be tested is the frontier itself. We may want to test

$$H_0 : m(X, Z) = m(X_1, Z), \quad \text{a.s.}, \tag{12}$$

where $X_1 \in \mathbb{R}^{p_1}$ is a subset of the input vector X. This means that the other inputs $X_2 \in \mathbb{R}^{p-p_1}$ are not relevant for the production frontier. More realistically, we may want to test if some of the inputs, say X_1 can be aggregated, say, $\xi_1 = a'X_1$, where a is a known vector (e.g., to allow the aggregation of inputs in different units)

$$H_0 : m(X, Z) = m(a'X_1, X_2, Z), \quad \text{a.s.}$$

For both cases the dependent variable is $G = \check{Y}$ and the quantity of interest $T(w)$ can be defined by[7]

$$T(w) = \mathbb{E}\left[f(X, Z)\big(\check{Y} - m(\xi, X_2, Z)\big)\mathbb{1}(W \leq w)\right].$$

An estimator of $T(w)$, analog to (10) is now given by

$$T_n(w) = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(X_i, Z_i)\big(\widehat{\check{Y}}_i - \widehat{m}(\xi_i, X_{2,i}, Z_i)\big)\mathbb{1}(W_i \leq w),$$

where $\widehat{m}$ is a nonparametric estimator of m based on the sample $\{(\widehat{\check{Y}}_i, \xi_i, X_{2,i}, Z_i)\}_{i=1}^n$. Then the bootstrap can be implemented similarly as in the preceding subsections to provide the p-value for this test.

[7]Similar expressions are available for the test (12), replacing $m(\xi, X_2, Z)$ by $m(X_1, Z)$.

Other restrictions can be tested following the same vein, like, e.g., partial linear models, or single index models.

4. Monte Carlo Simulations

In all the scenarios we will assume that the production technology is characterized by (1), where $X \sim \mathcal{U}[1,10]$, $Z \sim \mathcal{U}[0.1,1]$ and $U|x,z \sim \mathcal{N}_+(0,\sigma_U^2(x,z))$, where

$$\sigma_U(x,z) = \sigma_{U0} \times z_1^{b_{U1}} \times x_1^{b_{U2}}$$

and $V|x,z \sim \mathcal{N}(0,\sigma_V^2(x,z))$, where

$$\sigma_V(x,z) = \sigma_{V0} \times z_1^{b_{V1}} \times x_1^{b_{V2}}$$

and the production frontier is given by

$$m(x,z) = a_1 x_1 + a_2 \sin(x_1) + c_1 z_1. \tag{13}$$

While we tried many values for the parameters, the results presented below are for the case where $a_1 = 1$, $a_2 = 1$. This ensures that the technology satisfies free disposability, which is a typical assumption in production theory, though not necessary for our more general approach.

In all cases, for the regression in the first stage (to estimate r_1), we use least-squares cross-validation (LSCV) to estimate bandwidths for $W = (X,Z)$. Meanwhile, for the bandwidth used to construct the DGM test statistic we use $h = \delta n^{-1/(3q)}$, where q is the dimension of $W_1 = (X_1, Z_1)$ and δ is a tuning constant. This is in the spirit of Delgado & González Manteiga (2001).

4.1. **Results Without Separability.**

Prior to discussing some promising results for the DGM test we alert the reader to some issues we encountered when both x and z appear together (either in the frontier or in the inefficiency parameter). While no test is beyond reproach, we find these insights useful for empiricists adopting nonparametric methods to conduct stochastic frontier analysis and any subsequent inference based on it.

4.1.1. *Both x and z appear in inefficiency.* We assume that $V|x,z \sim \mathcal{N}(0,\sigma_V^2)$ (homoskedastic noise) with $\sigma_V = 1$ and that $U|x,z \sim \mathcal{N}_+(0,\sigma_U^2(x,z))$, where $\sigma_U(x,z) = \sigma_0 e^{\beta_1(z+x/10)}$ with $\beta_1 = 1$ and $\sigma_0 = 1$. The production frontier is specified as

$$m(x,z) = 2 + a_1 x + a_2 \sin(x).$$

The results for the DGM test presented below are for the case where $a_1 = 1$, we assess size by setting $a_2 = 0$ and power for $a_2 \in \{-1, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 1\}$; this ensures that the technology satisfies free disposability of inputs.

To speed up the computations for this illustration and to highlight an important practical issue, we follow the corresponding procedure. We generate samples of size n =2,500 for each value of a_2 in Equation (14). We then use LSCV to select the bandwidths for both the first and third stages of the SVKZ estimator with a Gaussian kernel deploying local-linear least-squares. Both of these regressions are estimated *including both* X and Z. The scale factors, i.e., c_ℓ from writing the bandwidth as $h_\ell = c_\ell \widehat{\sigma}_\ell n^{-1/6}$ for $\ell = X$ or Z with $\widehat{\sigma}_\ell$ the estimator of the standard deviation of the random variable ℓ, are stored and this process is repeated 50 times.[8] We then take the median of these scale factors to construct 'rule-of-thumb' bandwidths for our actual simulations. The aim here is to substantially reduce the computational burden and cut down on the additional noise that is introduced into the simulations through bandwidth selection when both x and z appear in either the conditional frontier or conditional inefficiency.

Meanwhile, for the test-related bandwidth we use $h = n^{-1/(3q)}$, where q is the dimension of (X, Z) over which we test for correct specification (i.e., $q = 1$) along with a second order Epanechnikov kernel. We conduct 499 Rademacher wild-residual bootstrap replications over 1,000 Monte Carlo trials for each sample size and value of a_2. In all of the simulations reported here, after $\widehat{\mu}_U(X_i, Z_i)$ is estimated, it is held fixed in the bootstrap replications.

Level accuracy plots for the DGM test are presented in the upper panels of Figures 1 and 2. These figures correspond to the case when $a_2 = 0$ and we assess the performance of the DGM test using either $\widehat{\widetilde{Y}}$ (which we term in the figure Y^*) or the corresponding Y value that results in treating conditional inefficiency as known.

As with the bivariate frontier size plots, the use of SVKZ to remove the component of the conditional mean of Y that is due to conditional inefficiency, results in severe size distortions. Treating $E[U|X = x, Z = z]$ as known instead of estimating it, produces an estimated size that is very near to nominal size across all sample sizes. That these size distortions appear is due to the nonparametric bias that stems from both X_i and Z_i appearing in the conditional mean of inefficiency, as discussed in Section 3.

Similar results hold for the assessment of power. Power plots for the DGM test are presented in the lower panels of Figures 1 and 2. The results are as expected in light of the

[8]The exponent 1/6 appears since the standard optimal rate for bandwidth decay is $1/(4 + q)$.

size findings. As a_2 increases the test does a better job detecting departures from correct parametric specification on X. This is true whether we test using $\widehat{\widehat{Y}}$ or the true values of the frontier though we need a caveat on the size distortions just discussed.

To discern if this performance was solely linked to nonparametric estimation of the model we also implemented the same process but used nonlinear least squares (i.e., parametric) to estimate the model. In this case we estimate our parameter vector θ by minimizing

$$\sum_{i=1}^{n}(Y_i - a_0 - a_1 X_i - a_2 \sin(X_i) + e^{\beta_1 X_i + \beta_2 Z_i})^2.$$

We then use $\sqrt{2/\pi}e^{\widehat{\beta}_1 X_i + \widehat{\beta}_2 Z_i}$ to correct our dependent variable prior to implementing the DGM test. In this case the NLS estimator for the model is correctly specified. Figure 3 presents the empirical size results from this exercise. As is evident, the size is quite poor, on par with the results when we used SVKZ to estimate the conditional mean of U prior to the correction of Y.

This suggests that the problem of size distortion is more general than the SVKZ per se, and is likely to be present when other semi- and nonparametric approaches are deployed, where by and large the inference frameworks are currently absent and so we hope it will be explored in future research endeavors.

4.1.2. *Both x and z appear in the frontier.* To further examine the performance of the DGM test when separability is not satisfied, we again assume that $V|x, z \sim \mathcal{N}(0, \sigma_V^2)$ (homoskedastic noise) with $\sigma_V = 1$ and that $U(x, z) \sim \mathcal{N}_+(0, \sigma_U^2(x, z))$, where $\sigma_U(x, z) = \sigma_0 e^{\beta_1 z}$ with $\beta_1 = 1$ and $\sigma_0 = 1$. The production frontier is specified as

$$m(x, z) = 2 + a_1(x + 10z) + a_2 \left(\sin(x) + \sin(10z)\right). \tag{14}$$

The results for the DGM specification test presented below are for the case where $a_1 = 1$. We assess size by setting $a_2 = 0$ and power for $a_2 \in \{-1, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 1\}$; this ensures that the technology satisfies free disposability and monotonicity of inputs. Again, to speed up the computations for this illustration we follow the same procedure as before.

Meanwhile, for the test-related bandwidth we use $h = n^{-1/(3q)}$, where q is the dimension of (X, Z) over which we test for correct specification (i.e., $q = 2$) along with a 4th order Epanechnikov kernel.[9] We conduct 499 Rademacher wild-residual bootstrap replications

[9]DGM suggested use of higher order kernels when the dimensionality of the covariates was greater than 1.

over 1,000 Monte Carlo trials for each sample size and value of a_2. In all of the simulations reported here, after $\widehat{\mu}_U(X_i, Z_i)$ is estimated, it is held fixed in the bootstrap replications.

Level accuracy plots for the DGM test are presented in the upper panels of Figures 4 to 5. These figures correspond to the case when $a_2 = 0$ and we assess the performance of the DGM test using either $\widehat{\tilde{Y}}$ (the original Y which has the mean subtracted off using the SVKZ approach) or Y with the true level of inefficiency subtracted off.[10]

The results are striking, the use of SVKZ to remove the component of the conditional mean of Y that is due to conditional inefficiency results in severe size distortions of the test (though, given our earlier results using the NLS estimator, it is likely a similar phenomena would arise with other nonparametric stochastic frontier estimators as well). Using the actual values of the frontier, i.e., treating $E[U|X = x, Z = z]$ as known instead of estimating it, produces estimated size that is very close to the nominal size, for any sample size.

Similar results hold for the assessment of power. The power plot for the DGM test are presented in the lower panels of Figures 4 and 5 for the same setups as detailed above. The results are as expected in light of the size findings. As a_2 increases the DGM test does a better job detecting departures from correct parametric specification on X. This is true whether we test using $\widehat{\tilde{Y}}$ or the true values of the frontier though, again, we need a caveat on the size distortions just discussed.

To conclude this section, it is worth emphasizing that without separability even a fully parametric stochastic frontier model estimated via NLS with all correct specifications, has shown similarly poor performance in terms of the size of the test (see Figure 3), even when the starting values in the optimization were the true values of the unknown parameters. Hence, expecting that the SVKZ estimator (and likely any other nonparametric estimator) would perform well in terms of size, when even correctly specified parametric approach cannot do so, appears to be unrealistic. In turn, this justifies the approach of Kim & Schmidt (2008) who only considered the case with separability in all their Monte Carlo scenarios and all theoretical derivations. As a result, the rest of the paper will also focus on the separable cases.

4.2. **Results With Separability.** Despite the poor performance of the DGM test to conduct inference on the correct specification of the stochastic frontier, the news is not all bad. Here we detail a set of simulations that demonstrate desirable performance of DGM when

[10]This allows us to assess the impact that bias stemming from nonparametric estimation of the first stage has on the performance of the DGM test.

we adopt the separability assumption, consistent with the testing environment used by Kim & Schmidt (2008) in their parametric framework.

4.2.1. *Significance Testing.* For the results presented here, we have $c_1 = 0$, $\sigma_{U0} = 0.5$; $b_{U1} \in \{0,\ 0.25,\ 0.5,\ 1,\ 1.5\}$ and $\sigma_{V0} = 0.2$; $b_{V1} = 0.5$, which makes the noise heteroskedastic, depending on Z. We also have $b_{U2} = 0.2$; $b_{V2} = 0.1$, i.e., with additional heteroskedasticity coming from X, influencing both V and U via their skedastic functions. We also tried many other values of parameters of the skedastic functions appearing right before Equation (13), including homoskedastic cases, and the results are similar with basically the same conclusions.

To accurately assess the performance of the DGM significance test we present level accuracy and power plots in the upper and lower panels of Figure 6, respectively. These curves are constructed using 1,000 Monte Carlo simulations with 499 Rademacher wild-residual bootstrap resamples within each simulation. Power is calculated assuming a size of 0.05. Overall, the results suggest that the estimated size (i.e., when $b_{u1} = 0$) is generally close to nominal size, except for fairly small samples $n = 50$ or 100. Meanwhile, the power increases as b_{u1} increases (for the same n) and increases in n (for the same b_{u1}), as is expected.

4.2.2. *Specification Testing.* In this section we detail a set of Monte Carlo simulations to investigate the performance of the test of correct parametric specification of the stochastic frontier detailed in Section 3.2. Here we consider sample sizes of $n \in \{25, 50, 100, 200, 400, 800\}$ for $X \sim U[1, 10]$ and $Z \sim U[0.1, 1]$. We assume that $V|x, z \sim \mathcal{N}(0, \sigma_v^2)$ (homoskedastic noise) with $\sigma_v = 1$ and that $U|x, z \sim \mathcal{N}_+(0, \sigma_U^2(x, z))$, where $\sigma_U(x, z) = \sigma_0 e^{\beta_1 z}$ with $\beta_1 = 1$ and $\sigma_0 = 1$.

The production frontier is specified as

$$m(x, z) = a_1 x + a_2 \sin(x).$$

The results presented below are for the case where $a_1 = 1$. We assess size by setting $a_2 = 0$ and power for $a_2 \in \{-1, -0.75, -0.5, -0.25, 0.25, 0.5, 0.75, 1\}$; again this ensures that the technology satisfies free disposability of inputs. Here we have the separable case with x only influencing output through the technology, and z only influencing output through inefficiency.

Both of these regressions are estimated *including both* X and Z. Meanwhile, for the test-related bandwidth we use $h = n^{-1/(3q)}$, where q is the dimension of (X, Z) over which we test for correct specification (i.e., here it is just X so $q = 1$) along with an Epanechnikov

kernel. We conduct 499 Rademacher wild-residual bootstrap replications over 1,000 Monte Carlo trials for each sample size and value of a_2. In all of the simulations reported here, after $\widehat{\mu}_U(X_i, Z_i)$ is estimated, it is held fixed in the bootstrap replications.

A level accuracy plot for the DGM test is presented in the upper panel of Figure 7. This figure corresponds to the case when $a_2 = 0$ and we assess the performance of the DGM test using $\widehat{\tilde{Y}}$ which is calculated by subtracting off the mean estimated through the SVKZ approach, which we term Y^* in the figure.

As we can see the DGM test displays only minor size distortions when using $\widehat{\tilde{Y}}$. This is most likely due to the fact that X and Z are separable and uncorrelated in this example coupled with the fact that there will be an inherent estimation error associated with the use of $\widehat{\tilde{Y}}$, which is made more pronounced with data-driven bandwidth selection. Moreover, keep in mind that in both stages of the SVKZ estimator, we are including both X and Z in the estimation and subsequent bandwidth selection; this undoubtedly will have an effect on performance.

Next we compare the power of the DGM test. The corresponding estimate power curve for the DGM test is presented in the lower panel of Figure 7. The results are as expected. As a_2 increases the test does a better job detecting departures from correct parametric specification on X. We see the classic $\top$ shape taking hold as n increases.

4.2.3. *Allowing correlation.* For the separable case, we also performed the same set of Monte Carlo simulations but allowed X and Z to be correlated to determine if this had any impact on the performance of the test using either observed output, Y or adjusted output $\widehat{\tilde{Y}}$. Here, rather than specify Z as uniform we instead generate Z as $\Phi(0.1X_i+\epsilon_i)$ where ϵ_i is distributed as $\mathcal{N}(0, 1)$. This puts the correlation of X and Z at ≈ 0.25.

The level accuracy plot for the DGM test is presented in the upper panel of Figure 8. As we can see the DGM test displays size distortions when using $\widehat{\tilde{Y}}$, of about 3%. This is due to the fact, for the $n = 800$ setting, across the 1,000 simulations, 33 trials produced a p-value exactly equal to 0. If we investigate these cases we see that they correspond to simulations where the SVKZ estimator had many instances of local "wrong skewness" (Simar & Wilson, 2010, 2021). As we are including both X and Z in the estimation and bandwidth selection, this undoubtedly makes it harder for the SVKZ estimator to detect structure.

If we remove the 0 p-value draws, we see that the DGM test works correctly when using $\widehat{\tilde{Y}}$, see Figure 9. These results suggest that practitioners deploying the DGM test may wish

to purge 0 p-value draws in the sampling process as they are likely to distort the true nature of the implications of the test.

Next we compare the power of the DGM test when X and Z are correlated. Estimated power curves for the DGM test are presented in the lower panel of Figure 8 with the results as expected. As a_2 increases the test does a better job detecting departures from correct parametric specification on X.

We see from these simulations that one can test accurately for correct functional form when separability holds. Moreover, correlation between X and Z did not appear to adversely impact the performance of the test beyond the issue of local wrong skewness, which is easy for a practitioner to diagnose. Thus, consistent with the findings of Kim & Schmidt (2008), one can confidently assess the parametric structure of a production frontier in a nearly nonparametric setting.

5. Conclusions

Consistent specification testing is a bedrock principle in all areas of applied statistical research. While a variety of tests for correct specification have been proposed for standard regression settings, practically none have been thoroughly scrutinized in the realm of stochastic frontiers. We have tried to fill this gap here, proposing a robust omnibus testing facility based on the insights of DGM. Modifications were needed as direct application of DGM does not follow when one is dealing with a frontier.

Our simulations suggest that when separability holds between those covariates that impact the location of the frontier and the level of inefficiency, then the DGM test works reasonably well both for testing the statistical significance of the impact of covariates on conditional efficiency, as well as for correct parametric specification. The significance results can be viewed as the nonparametric generalization of the work of Kim & Schmidt (2008) for the parametric setting.

We also discovered some limitations of the tests that apparently were not perceived in the earlier literature. Specifically, when the assumption of separability is violated, the DGM test can suffer size distortions, which also lead to compromised power. This stems from the manner in which bias manifests in the absence of separability and if it is present for other estimators and other tests. Future work to examine how to mitigate this bias for the purpose of inference are ongoing. We also note that there potentially exist further generalizations to accommodate panel data and to adapt developments in Belotti & Ilardi

(2018) to the nonparametric setting as well, which are left for future research. Our overall general conclusion is in fact similar to the one reached by Simar & Wilson (2007) for the competing alternative (DEA+truncated regression), namely: Inference for the SFM in the semi- and nonparametric context is quite challenging, yet possible under the separability assumption along with an appropriate bootstrap.

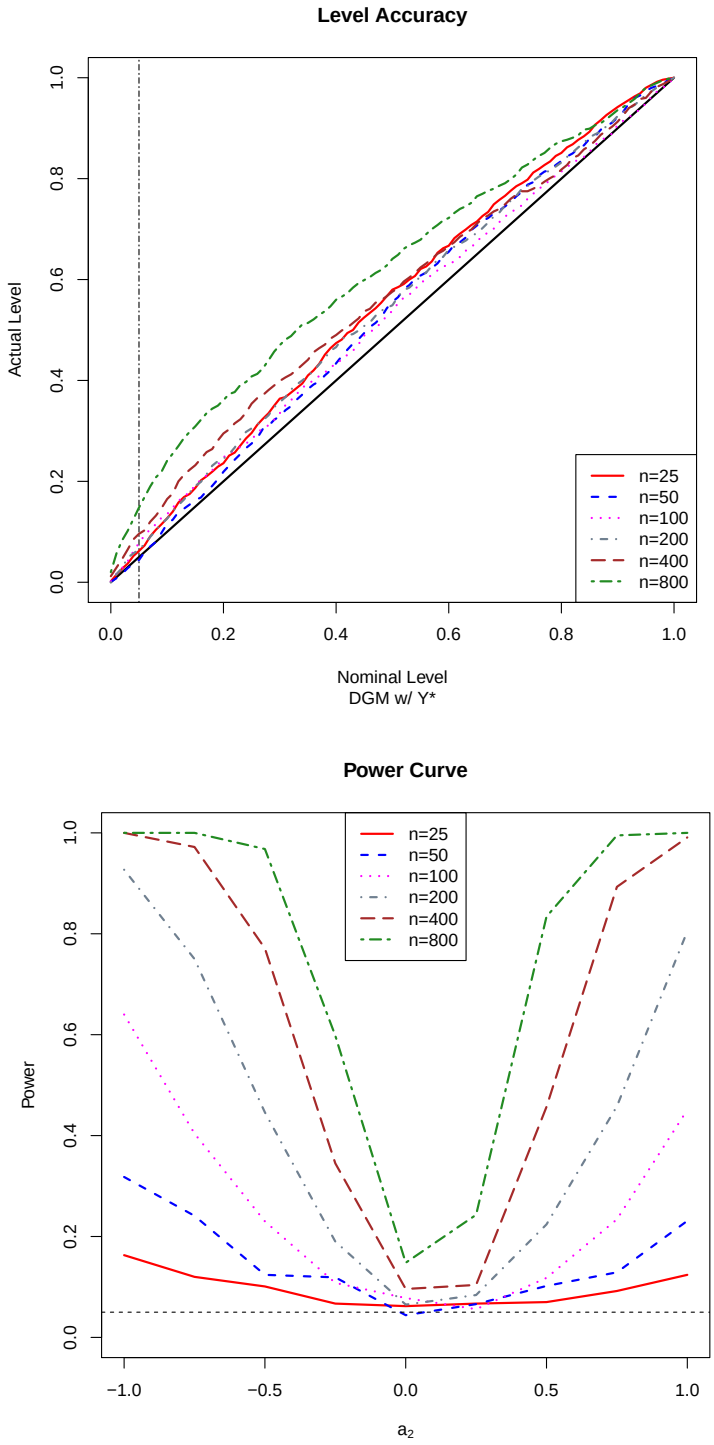


FIGURE 1. Level accuracy and power plots for DGM specification test using corrected output, $\widehat{\tilde{Y}}$ (Y^* in the figure), when both X and Z belong to the inefficiency.

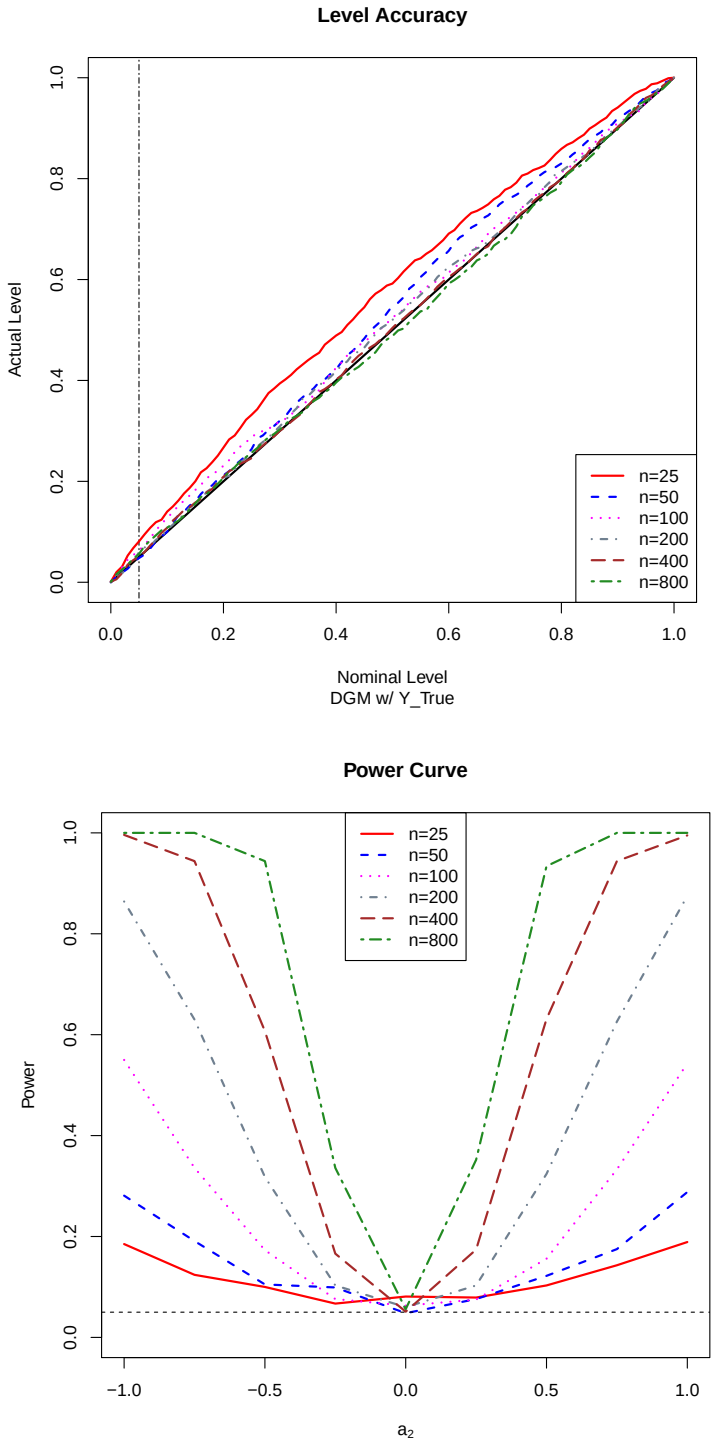


FIGURE 2. Level accuracy and power plots for DGM specification test using true frontier when both X and Z belong to the inefficiency.

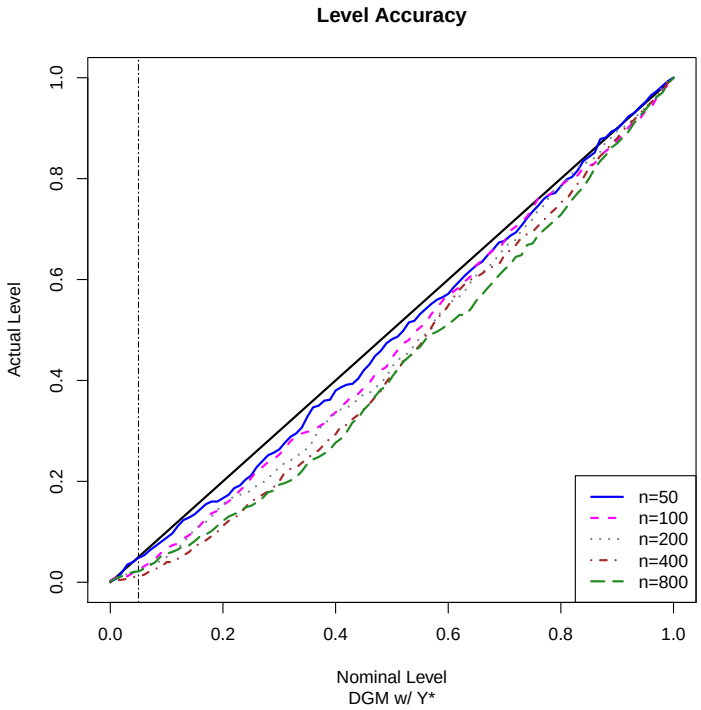


FIGURE 3. Level accuracy for DGM specification test using parametrically estimated model via NLS when both X and Z belong to the inefficiency.

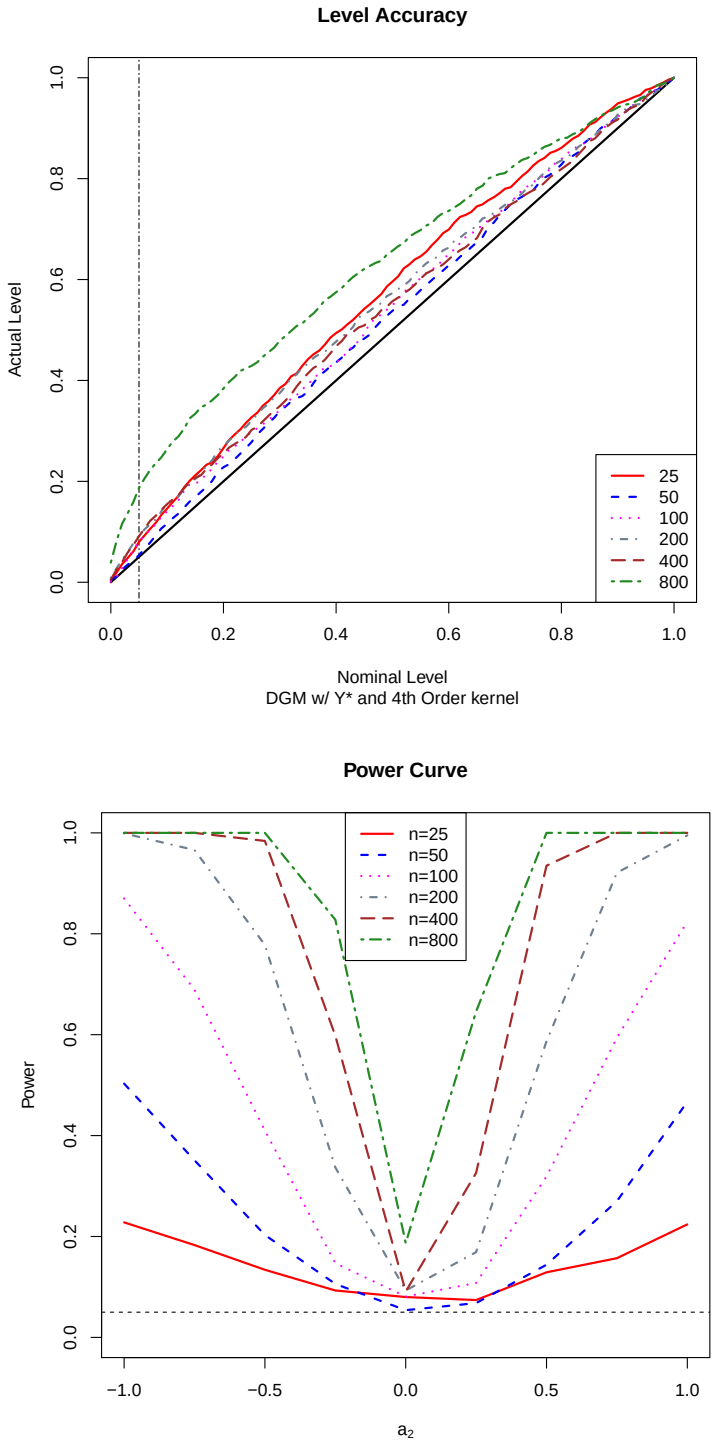


FIGURE 4. Level accuracy and power plots for DGM specification test using corrected output, $\widehat{\widehat{Y}}$, when both X and Z belong in the frontier.

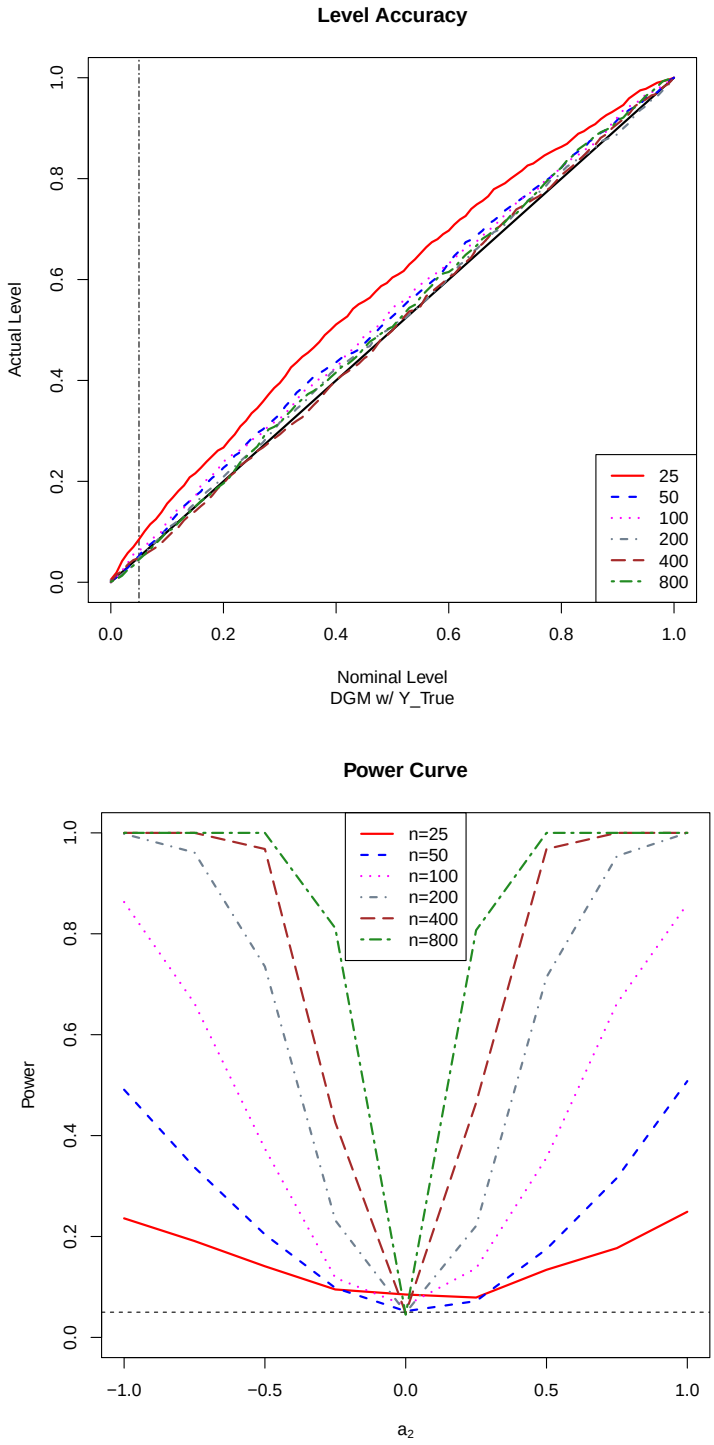


FIGURE 5. Level accuracy and power plots for DGM specification test using true output when both X and Z belong in the frontier.

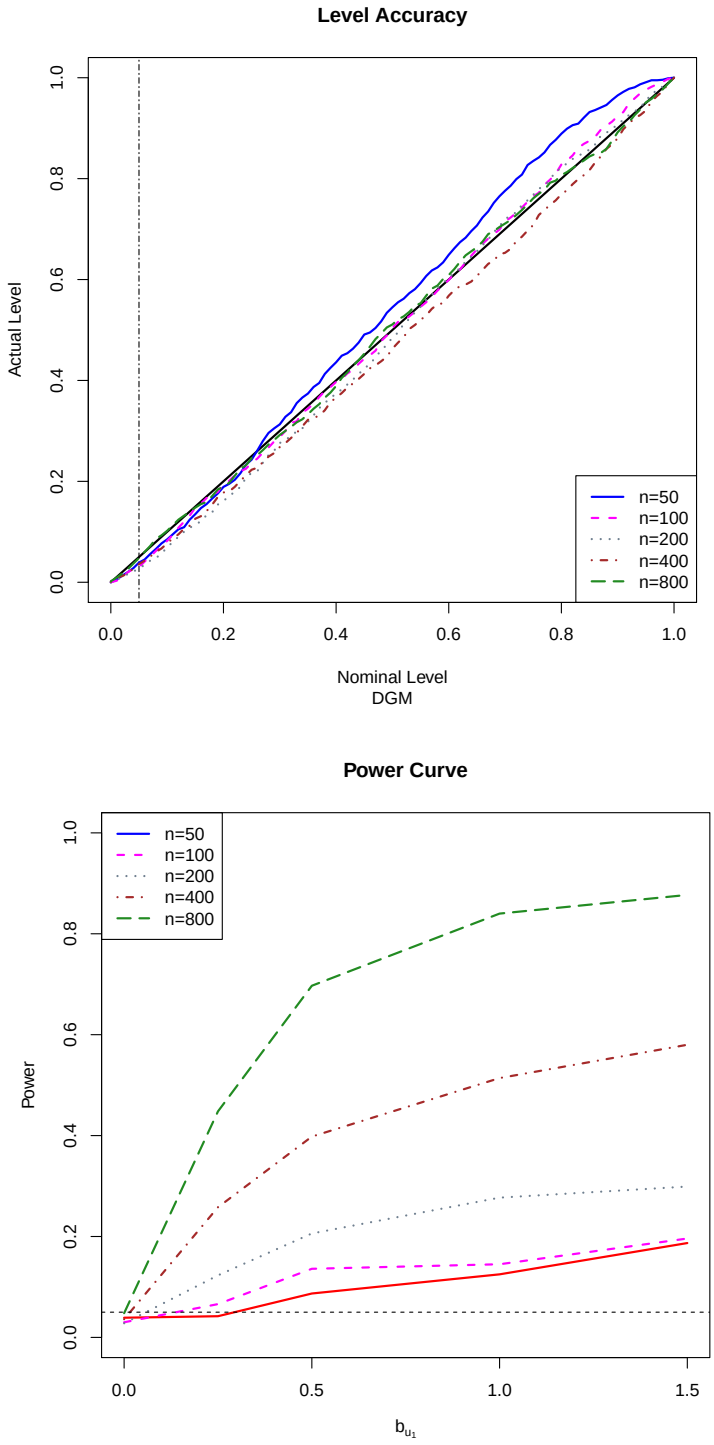


FIGURE 6. Level accuracy and power plots for DGM significance test under separability and using corrected output Y^*.

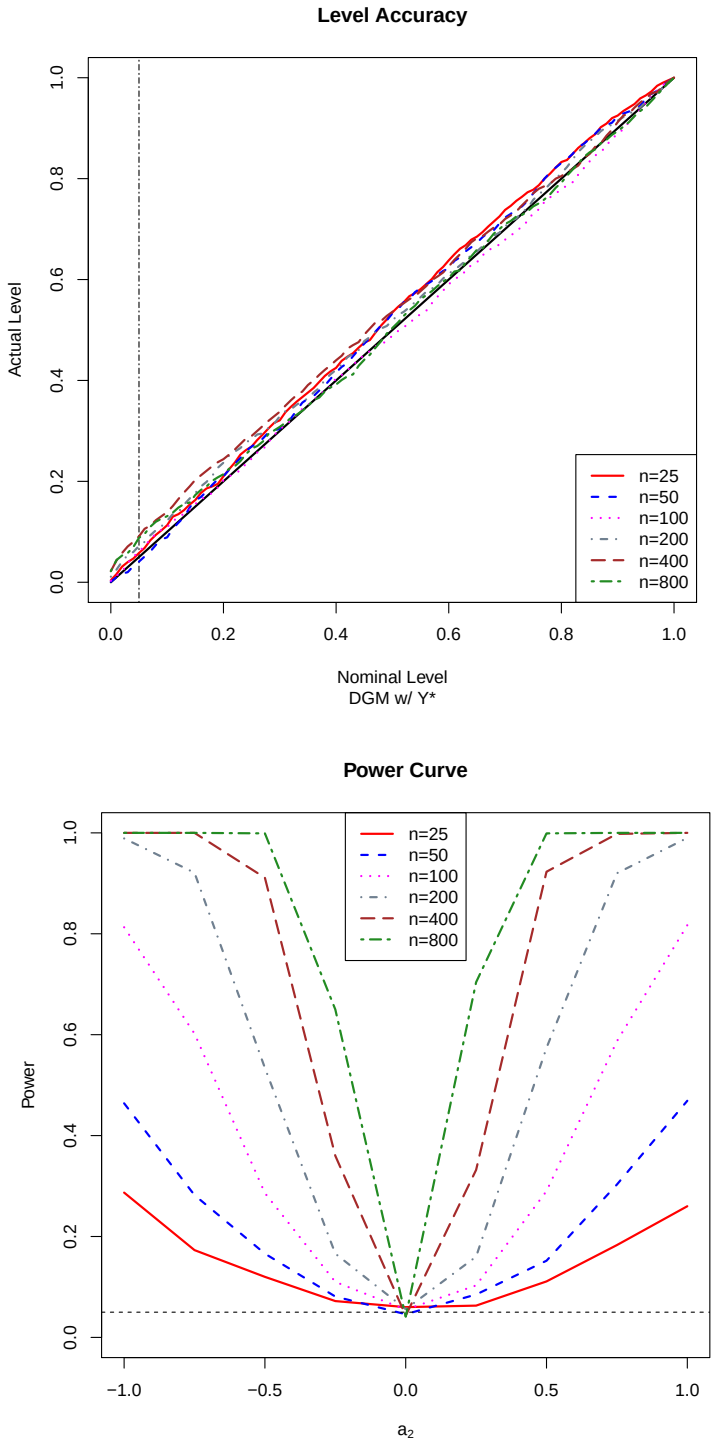


FIGURE 7. Level accuracy and power plots for DGM specification test using corrected output, $\widehat{\widehat{Y}}$ under separability.

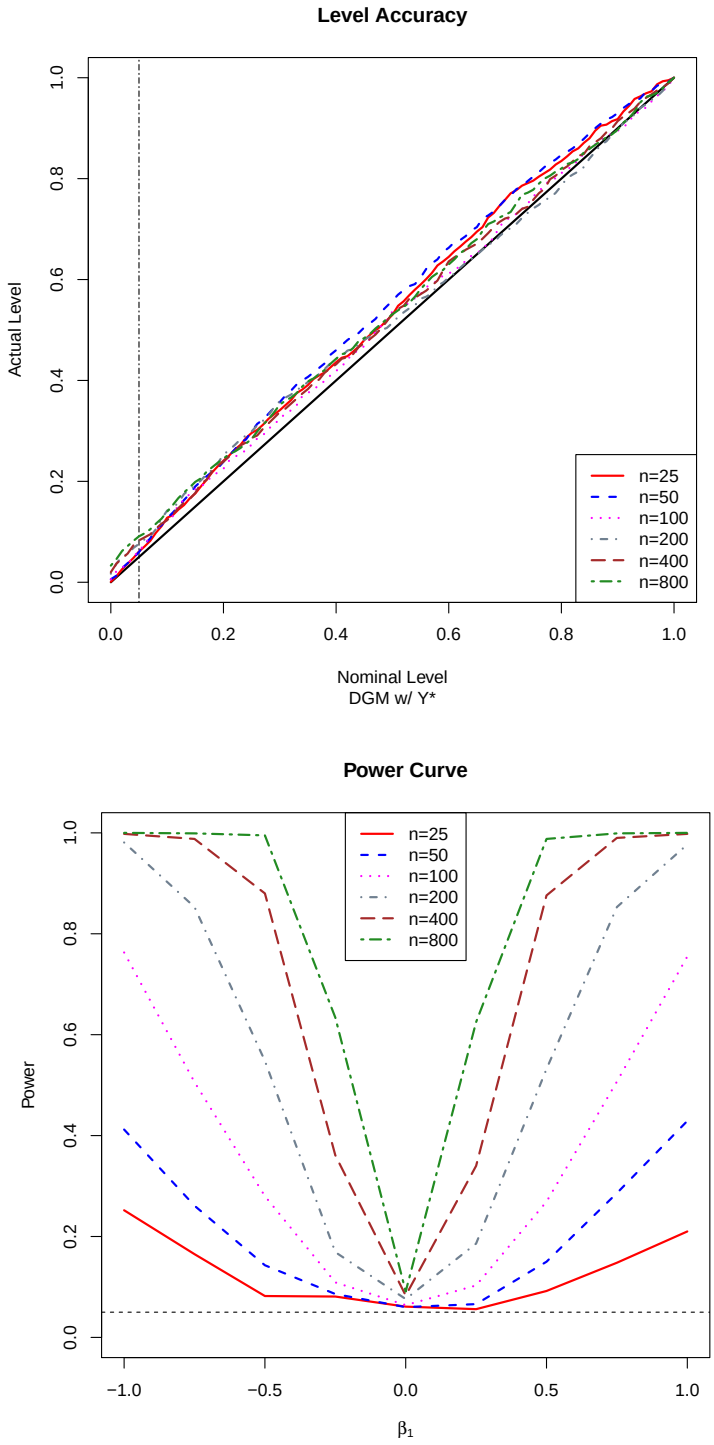
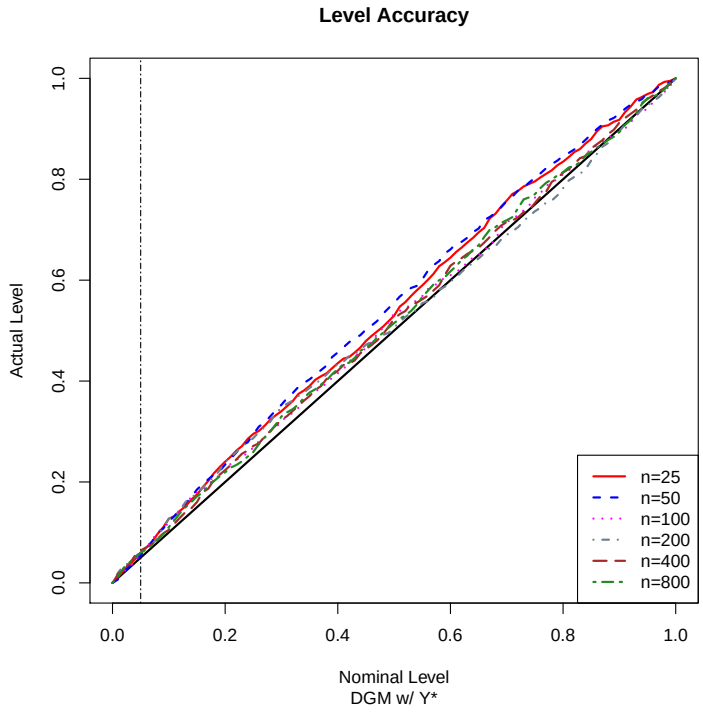


FIGURE 8. Level accuracy and power plots for DGM specification test using corrected output, $\widehat{\widehat{Y}}$, X and Z correlated.

FIGURE 9. Level accuracy plots for DGM specification test using corrected output, $\widehat{\widehat{Y}}$, removing 0 p-values.



References

Amsler, C., Prokhorov, A. & Schmidt, P. (2016), 'Endogeneity in stochastic frontier models', *Journal of Econometrics* **190**, 280–288.

Amsler, C., Prokhorov, A. & Schmidt, P. (2017), 'Endogeneity environmental variables in stochastic frontier models', *Journal of Econometrics* **199**, 131–140.

Behr, A. (2010), 'Quantile regression for robust bank efficiency score estimation', *European Journal of Operational Research* **200**, 568–581.

Belotti, F. & Ilardi, G. (2018), 'Consistent inference in fixed-effects stochastic frontier models', *Journal of Econometrics* **202**(2), 161–177.

Chen, Y.-T., Hsu, Y.-C. & Wang, H.-J. (2020), 'A Stochastic Frontier Model with Endogenous Treatment Status and Mediator', *Journal of Business & Economic Statistics* **38**(2), 243–256.

Chen, Y.-Y., Schmidt, P. & Wang, H.-J. (2014), 'Consistent estimation of the fixed effects stochastic frontier model', *Journal of Econometrics* **181**(1), 65–76.

Delgado, M. A. & González Manteiga, W. (2001), 'Significance testing in nonparametric regression based on the bootstrap', *Annals of Statistics* **29**(5), 1469–1507.

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and its Application*, Chapman and Hall.

Fan, J., Zhang, C. & Zhang, J. (2001), 'Generalized likelihood ratio statistics and wilks phenomenon', *Annals of Statistics* **29**(1), 153–193.

Fan, Y. & Li, Q. (1996), 'Consistent model specification tests: Omitted variables and semiparametric functional forms', *Econometrica* **64**(4), 865–890.

Fan, Y., Li, Q. & Weersink, A. (1996), 'Semiparametric estimation of stochastic production frontier models', *Journal of Business & Economic Statistics* **14**(4), 460–468.

Glass, A. J., Kenjegalieva, K. & Sickles, R. C. (2016), 'A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers', *Journal of Econometrics* **190**(2), 289–300.

Greene, W. H. (2005), 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics* **126**(2), 269–303.

Hafner, C., Manner, H. & Simar, L. (2018), 'The "wrong skewness" problem in stochastic frontier models: A new approach', *Econometric Reviews* **37**, 380–400.

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

Henderson, D. J. & Parmeter, C. F. (2015), *Applied Nonparametric Econometrics*, Cambridge University Press, Cambridge, Great Britain.

Horrace, W. C. & Wright, I. A. (2020), 'Stationary points for parametric stochastic frontier models', *Journal of Business & Economic Statistics* **38**, 516–526.

Kim, M. & Schmidt, P. (2008), 'Valid test of whether technical inefficiency depends on firm characteristics', *Journal of Econometrics* **144**(2), 409–427.

Kneip, A., Simar, L. & Van Keilegom, I. (2015), 'Frontier estimation in the presence of measurement error with unknown variance', *Journal of Econometrics* **184**, 379–393.

Kumbhakar, S. C., Park, B. U., Simar, L. & Tsionas, E. G. (2007), 'Nonparametric stochastic frontiers: A local maximum likelihood approach', *Journal of Econometrics* **137**(1), 1–27.

Kutlu, L., Tran, K. C. & Tsionas, M. G. (2019), 'A time-varying true individual effects model with endogenous regressors', *Journal of Econometrics* **211**(2), 539–559.

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

Li, Q. & Wang, S. (1998), 'A simple consistent bootstrap test for a parametric regression function', *Journal of Econometrics* **87**(2), 145–165.

Martins-Filho, C. B. & Yao, F. (2015), 'Semiparametric stochastic frontier estimation via profile likelihood', *Econometric Reviews* **34**, 413–451.

Olson, J. A., Schmidt, P. & Waldman, D. A. (1980), 'A Monte Carlo study of estimators of stochastic frontier production functions', *Journal of Econometrics* **13**, 67–82.

Orea, L. & Álvarez, I. C. (2019), 'A new stochastic frontier model with cross-sectional effects in both noise and inefficiency terms', *Journal of Econometrics* **213**(2), 556–577.

Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, New York.

Park, B. U., Simar, L. & Zelenyuk, V. (2015), 'Categorical data in local maximum likelihood: theory and applications to productivity analysis', *Journal of Productivity Analysis* **43**(1), 199–214.

Parmeter, C. F. & Zelenyuk, V. (2019), 'Combining the virtues of stochastic frontier and data envelopment analysis', *Operations Research* **67**(6), 1628–1658.

Simar, L., Van Keilegom, I. & Zelenyuk, V. (2017), 'Nonparametric least squares methods for stochastic frontier models', *Journal of Productivity Analysis* **47**(3), 189–204.

Simar, L. & Wilson, P. W. (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics* **136**(1), 31–64.

Simar, L. & Wilson, P. W. (2010), 'Inferences from cross-sectional, stochastic frontier models', *Econometric Reviews* **29**(1), 62–98.

Simar, L. & Wilson, P. W. (2021), Nonparametric, stochastic frontier models with multiple inputs and outputs. LIDAM Discussion Paper ISBA; 2021/03.

Ullah, A. (1985), 'Specification analysis and econometric models', *Journal of Quantitative Economics* **2**, 187–209.

Wang, H.-J. & Ho, C.-W. (2010), 'Estimating fixed-effect panel stochastic frontier models by model transformation', *Journal of Econometrics* **157**(2), 286–296.

Zheng, J. X. (1996), 'A consistent test of functional form via nonparametric estimation techniques', *Journal of Econometrics* **75**(2), 263–289.