Semi-Nonparametric Identification and Estimation of the Stochastic Frontier Model

Herman J. Bierens*
Pennsylvania State University, USA

Hung-Pin Lai
National Chung Cheng University, Taiwan

May 9, 2019

Abstract
To be done.

1 Introduction

Since the pioneering works of Aigner et al. (1977, ALS hereafter) and Meesuen and van den Broeck (1977, MB hereafter), the stochastic frontier (SF) model has been widely applied in empirical studies of production inefficiency. The main characteristic of the SF model is that the composite error of the model contains two random components. One is the stochastic error \( V \), which is usually assumed as zero mean normally distributed, and the other one is the nonnegative random component \( W \) that captures the production inefficiency in the model. In particular, the linear SF model takes the form

\[
Y = \alpha_0 + X'\beta_0 + V - W, \tag{1}
\]

where \( V \) is an error term with the usual properties, i.e., \( E[V] = 0, E[V^2] = \sigma_0^2 < \infty \), \( X \) is the vector of regressors, and \( W \) is an unobserved nonnegative random variable measuring the distance from the production frontier.

*Professor Emeritus of Economics. Since I no longer have an office at PennState, please address correspondence by email only to hbierens@psu.edu.
For example, suppose that the original model is a Cobb-Douglas production function, i.e.,

\[ Q = \exp(V) \exp(-W) \exp(\alpha_0)K^{\beta_{0,1}}L^{\beta_{0,2}}, \]  

so that

\[ \ln(Q) = \alpha_0 + \beta_{0,1} \ln(K) + \beta_{0,2} \ln(L) + V - W, \]  

(2)

where \( \overline{Q} = \exp(V) \exp(\alpha_0)K^{\beta_{0,1}}L^{\beta_{0,2}} \) is the production frontier of an individual firm, \( \exp(-W) = Q/\overline{Q} \) represents the relative distance from the production frontier, and \( V \) is a regression error satisfying the standard regression condition \( E[V|X,W] = 0 \) a.s., interpreting \( W \) as an unobserved explanatory variable.

In order to estimate this model by parametric maximum likelihood (ML), one needs to derive the distribution of the composite error \( V - W \) under appropriate distribution assumptions of \( V \) and \( W \). The most common assumptions imposed on \( W \) include the half normal (ALS, 1977), exponential (MB, 1977) and truncated normal (Stevenson, 1980) distributions. Some other less frequently used distributions for \( W \) include the gamma distribution (Greene, 1980a,b and Stevenson, 1980), the four parameter Pearson density (Lee, 1983), the uniform distribution (Li, 1996), the binomial distribution (Carree, 2002), the beta distribution (Gagnepain and Ivaldi, 2002) and the double truncated normal distribution (Almanidis et al., 2014). Baccouche and Kouki (2003) conducted a sensitivity analysis for the most commonly used distributions (half normal, truncated normal and exponential distributions) and show that the estimates of technical efficiency depend heavily on the assumed distribution. Therefore, one criticism of the SF model is that it is too heavily parameterized and that it is likely that the misspecified inefficiency distribution will impact the estimates and further inferences based on them. The key empirical issue to implement the parametric ML approach is how to choose a correct distribution for \( W \), and then derive the density for the composite error. However, not all distributional assumptions will provide closed form solutions for the composite error distribution.

The main objective of this paper is to propose a semi-nonparametric (SNP)\(^1\) sieve estimation approach to estimate the SF model, where the density of \( W \) is modeled semi-nonparametrically, similar to Bierens (2014a,b). To the best of our knowledge, this has not yet been done in the literature.

\(^1\)Semi-nonparametric (SNP) models are models where only a part of the model is parametrized, and the non-specified part is an unknown function which is represented by an infinite series expansion. Therefore, SNP models are in essence models with infinitely many parameters.
The existing estimation approaches for semi-parametric \(^2\) SF (SP-SF hereafter) models are limited. Fan et al. (1996, FLW hereafter) lessen the dependence on parametric specification on the production frontier by replacing the density of \(W\) by its kernel estimator and then estimate the model by an ML approach. Ferrara and Vidoli (2017) suggest using an additive model for the frontier specification under the same framework as FLW. Earlier, Martin-Filho and Yao (2011) further extended the FLW’s approach by local likelihood and maximizing the profile likelihood.

Although the normality of \(V\) is widely accepted, and so do we, the distribution assumption on \(W\) has raised some debates in the literature. Some focused on using different distributions of \(W\) and others \(^3\) focused on hypothesis testing of the inefficiency distributions. This motivates the second type of SP-SF models that assumes a parametric frontier but leaves the distribution of inefficiency unspecified. For instance, see Winsten (1957), Afriat (1972) and Horrace and Parmeter (2011). Both Winsten (1957) and Afriat (1972) estimated the frontier equation by ordinary least squares (OLS) and then corrected/modified the biased OLS intercept through shifting it up towards an estimate of \(E[W]\). The corrected/modified OLS residuals are then used to provide estimates of inefficiency. In the same manner, Horrace and Parmeter (2011) also estimated the frontier equation by OLS. Given the residuals and the normality assumption for \(V\), they then applied the deconvolution technique proposed by Meister (2006) and Hall and Simar (2002) to identify the intercept of the SF equation and estimate the density of the inefficiency. The common characteristic of these methods is that they all used two-step procedures and that the prediction of the individual-specific inefficiency was not based on the conditional expectation \(E[W|V-W]\), but on the unconditional expectation \(E[W]\). Consequently, the information regarding the composite error is not efficiently used for predicting individual-specific inefficiency. How to predict the inefficiency as in Jondrow et al. (1982, JLMS) is still an unsolved issue in the existing SP-SF literature, but will be solved in the present paper.

\(^2\)Semi-parametric (SP) models have in common with SNP models that only a part of the model is parametrized, and the non-specified part is an unknown function. However, in SP models this unknown function is usually modeled nonparametrically, for example by using kernel estimators or related nonparametric estimators, given the parameters of the specified part of the model.

\(^3\)For instance, see Wang et al. (2011), Chen and Wang (2012), Hajargasht and Griffiths (2018).
1.1 Overview of the main results

In this paper, we focus on estimating the second type of SF model. Instead of dealing with the deconvolution problem as in the existing methods, we take a bottom-up approach by modeling the density of $W$ seminonparametrically, while maintaining the usual conditions that $V \sim \mathcal{N}(0, \sigma_0^2)$ and $X$, $V$ and $W$ are independent. On the basis of the OLS residuals and the OLS estimate of the regression intercept $\alpha_0 - E[W]$ we can strongly consistent estimate the pseudo-Laplace transform $E[\exp(-t.(W - V - \alpha_0))] = \exp(\alpha_0 t) \cdot \exp(\sigma_0^2 t^2/2) \cdot E[\exp(-t.W)]$, $t > 0$, where $\exp(\sigma_0^2 t^2/2)$ is the well-known moment generating function of $V \sim \mathcal{N}(0, \sigma_0^2)$ and $E[\exp(-t.W)]$ is the actual Laplace transform of the inefficiency variable $W$. In particular, it will be shown how to construct an empirical process $\Upsilon_N(t)$, say, on the basis of a random sample $\{(Y_i, X_i)\}_{i=1}^N$, such that for any constant $c > 0$,

$$\sup_{t \in [0,c]} |\Upsilon_N(t) - \alpha_0 t - \sigma_0^2 t^2/2 - \ln (E[\exp(-t.W)])| \overset{a.s.}{\to} 0$$  

(3) as $N \to \infty$.\(^4\)

Moreover, any density $f(w)$ on $[0, \infty)$ can be modeled seminonparametrically (in various ways) as $f(w|\delta)$, say, where $\delta = \{\delta_m\}_{m=1}^\infty$ is an infinite-dimensional parameter satisfying $\sum_{m=1}^\infty \delta_m^2 < \infty$, and the actual density $f_0(w)$ of $W$ corresponds to the SNP specification $f(w|\delta^0)$, say. These results give rise to an empirical objective function of the form

$$Q_N(\alpha, \sigma^2, \delta|c) = \int_0^c (\Upsilon_N(t) - \alpha t - \sigma^2 t^2/2 - \ln \left( \int_0^\infty \exp(-t.w) f(w|\delta) dw \right) )^2 dt$$  

(4)

for given constant $c > 0$.

Admittedly, at first sight the exact computation of the integrals in (4) does not look feasible. However, the Laplace transform involved (i.e., the integral in the log) can be computed exactly for a particular class of SNP densities $f(w|\delta)$, and then the integral over $[0, c]$ can be estimated strongly consistent by Monte Carlo integration, where most of the heavy computations can be done in advance, regardless the parameter values.

Given $\delta$, we propose to first concentrate $\alpha$ and $\sigma^2$ out by minimizing $Q_N(\alpha, \sigma^2, \delta|c)$ to $\alpha$ and $\sigma^2$, which yield closed form expressions for the resulting solutions $\alpha_N(\delta)$ and $\sigma^2_N(\delta)$, respectively. Next, minimizing the concentrated objective function $\tilde{Q}_N(\delta|c) = Q_N(\alpha_N(\delta), \sigma^2_N(\delta), \delta|c)$ via a sieve

\(^4\)Throughout this paper "a.s." stands for "almost surely", meaning that the property involved, in particular convergence, holds with probability 1.
estimation approach results in a consistent sieve estimator \( \hat{\delta}_N \) of the infinite-dimensional true parameter \( \delta^0 \), in the sense that \( p \lim_{N \to \infty} ||\hat{\delta}_N - \delta^0|| = 0 \), so that \( f(w|\hat{\delta}_N) \) is a consistent estimate of the density \( f_0(w) \) of \( W \). Moreover, we then also have \( p \lim_{N \to \infty} \alpha_N(\hat{\delta}_N|c) = \alpha_0 \) and \( p \lim_{N \to \infty} \sigma^2_N(\hat{\delta}_N|c) = \sigma^2_0 \).

In the fully parametric SF model the identification of the parameters of the distributions of \( V \) and \( W \) is usually not an issue. However, due to the flexibility of the SNP specification \( f(w|\delta) \) of the density of \( W \), semi-nonparametric identification of the SF model is a non-trivial issue. The SNP identification issue will be addressed in two ways. First, it will be shown that under mild conditions the model is semi-nonparametrically identified regardless the type of the distribution of \( W \), thus allowing for discrete, continuous and/or mixed discrete-continuous distributions. Second, in the case that \( W \) has an absolutely continuous distribution with continuous density it will be shown that pointwise in \( \delta \), \( Q_N(\delta|c) \overset{a.s.}{\to} Q(\delta|c) \) and that \( Q(\delta|c) = 0 \) if and only if \( \delta = \delta^0 \).

### 1.2 Plan of the paper

The plan of the paper is as follows. In section 2 we summarize the conditions for the strong consistency of the OLS estimator of \( \beta_0 \). In section 3 we show how to estimate the pseudo-Laplace transform of \( W - V - \alpha_0 \) consistently. On the basis of the latter result we show in section 4 that under a regularity condition the constant \( \alpha_0 \), the variance \( \sigma^2_0 \) of \( V \) and the distribution of \( W \) are identified, regardless the type of the distribution of \( W \). In section 5 we show how to model the density of \( W \) semi-nonparametrically, using the results in Bierens (2014a). In section 6 derive the asymptotic properties of the objective function (4) and its concentrated version \( Q_N(\delta|c) = Q_N(\alpha_N(\delta), \sigma^2_N(\delta), \delta|c) \) including the properties of \( \alpha_N(\delta) \) and \( \sigma^2_N(\delta) \). In section 7 we show that the parameter \( \delta^0 \) in the SNP specification \( f(w|\delta^0) \) of the actual density \( f_0(w) \) of \( W \) is identified, so that \( f_0(w) \) itself is identified. In section 8 we show that the sieve estimator \( \hat{\delta}_N \) on the basis of the concentrated objective function \( Q_N(\delta|c) \) is consistent, and \( \hat{\alpha}_N = \alpha_N(\hat{\delta}_N) \overset{p}{\to} \alpha_0 \), \( \hat{\sigma}^2_N = \sigma^2_N(\hat{\delta}_N) \overset{p}{\to} \sigma^2_0 \).

In section 9 we show how to estimate the technical efficiency (TE) index \( E[\exp(-W_i)|V_i - W_i] \) for each firm \( i \), and the related TE function \( E[\exp(-W)|V - W = u] \), \( u \in \mathbb{R} \). In section 10 we plan to present the results of a numerical

---

5 The norm \( ||\cdot|| \) involved is defined as \( ||\delta|| = \sqrt{\sum_{m=1}^\infty \delta_m^2} \) for \( \delta = \{\delta_m\}_m \).

6 In the sense that \( p \lim_{N \to \infty} \int_0^\infty |f(w|\hat{\delta}_N) - f_0(w)|dw = 0 \).
experiment, and in section 11 we plan to apply our SNP approach to a wage frontier model, using Taiwanese data. Finally, section 12 is an appendix containing the proofs of two lemmas and a theorem in this paper, and a subsection on how to compute SNP Laplace transforms exactly.

2 Estimation of $\beta_0$

The following conditions on the model variables are standard:

**Assumption 1.**
(a) The model variables $X_i$, $V_i$ and $W_i$ are independent of each other, and $(X_i, V_i, W_i)$ is i.i.d. as $(X, V, W)$ across observations. The variables $Y_i$ and $X_i$ are observed for $i = 1, 2, \ldots, N$.
(b) $X \in \mathbb{R}^p$, $E[X^'X] < \infty$ and $\Sigma = \text{Var}(X)$ is nonsingular.\(^7\)
(c) $V$ is distributed as $\mathcal{N}(0, \sigma_0^2)$.
(d) $\text{Pr}[W < 0] = 0$ and $E[W^2] < \infty$,
(e) $\text{Pr}[W \leq w] > 0$ for all $w > 0$.

Now write the SF model (1) as

$$Y_i = \alpha_0 - E[W] + X_i' \beta_0 + V_i - (W_i - E[W]) = \mu_0 + X_i' \beta_0 + U_i, \quad i = 1, 2, \ldots, N,$$

where $\mu_0 = \alpha_0 - E[W]$ and $U_i = V_i - W_i + E[W]$. Since by Assumption 1, $E[U_i|X_i] = E[U_i] = 0$ and $E[U_i^2] = \sigma_{\mu}^2 + \sigma_{W}^2 < \infty$, where $\sigma_{W}^2$ is the variance of $W$, the parameters $\mu_0$ and $\beta_0$ in model (5) can be estimated strongly consistent and asymptotic normally by OLS. In particular, as is well-known, under Assumption 1 the OLS estimator $\hat{\beta}_N$ of $\beta_0$ satisfies

$$E[\hat{\beta}_N] = \beta_0, \quad \hat{\beta}_N \xrightarrow{a.s.} \beta_0, \quad \text{and} \quad \sqrt{N}(\hat{\beta}_N - \beta_0) \xrightarrow{d} \mathcal{N}_p \left(0, (\sigma_{\mu}^2 + \sigma_{W}^2)\Sigma^{-1}\right)$$

as $N \to \infty$, and the OLS estimator $\hat{\mu}_N = (1/N)\sum_{i=1}^N (Y_i - \hat{\beta}_N X_i)$ of $\mu_0$ satisfies

$$E[\hat{\mu}_N] = \mu_0 = \alpha_0 - E[W], \quad \hat{\mu}_N \xrightarrow{a.s.} \mu_0 = \alpha_0 - E[W]$$

\(^7\)The latter excludes that one of the components of $X$ is a constant, as motivated by the Cobb-Douglas example (2).
as $N \to \infty$. Moreover,

$$\hat{s}_N^2 = \frac{1}{N} \sum_{i=1}^N \left( Y_i - \hat{\beta}_N X_i - \hat{\mu}_N \right)^2 \overset{a.s.}{\Rightarrow} \sigma_0^2 + \sigma_W^2 = \sigma_0^2 + E[W^2] - (E[W])^2. \quad (8)$$

These results trivially imply that

$$\lim_{N \to \infty} \Pr[\alpha_0 \geq \hat{\mu}_N] = 1, \quad \lim_{N \to \infty} \Pr[0 < \sigma_0^2 \leq \hat{s}_N^2] = 1. \quad (9)$$

### 3 Estimation of the pseudo-Laplace transform of $W - V - \alpha_0$

Recall from (1) that

$$X_i' \hat{\beta}_N - Y_i + ||X_i|| = X_i' (\hat{\beta}_N - \beta_0) + ||X_i|| + W_i - V_i - \alpha_0. \quad (10)$$

The reason for adding the term $||X_i||$ to both sides of this equation is that by (6),

$$X_i' (\hat{\beta}_N - \beta_0) + ||X_i|| \geq (1 - ||\hat{\beta}_N - \beta_0||)||X_i|| \geq 0 \text{ a.s.}$$

for sufficient large $N$. In particular, (6) implies that there exists an $N_0 \in \mathbb{N}$ such that $\Pr[||\hat{\beta}_N - \beta_0|| \leq 1] = 1$ whenever $N \geq N_0$. In the latter case we have, for any constant $c > 0$,

$$\sup_{t \in [0,c]} \left| \frac{1}{N} \sum_{i=1}^N \exp \left( -t. \left( X_i' \hat{\beta}_N - Y_i + ||X_i|| \right) \right) - E \left[ \exp \left( -t. \left( X' \xi + ||X|| + W - V - \alpha_0 \right) \right) \right]_{\xi = \hat{\beta}_N - \beta_0} \right| \leq \sup_{||\xi|| \leq 1, t \in [0,c]} \left| \frac{1}{N} \sum_{i=1}^N \exp \left( -t. \left( X_i' \xi + ||X_i|| + W_i - V_i - \alpha_0 \right) \right) - E \left[ \exp \left( -t. \left( X' \xi + ||X|| + W - V - \alpha_0 \right) \right) \right]_{\xi = \hat{\beta}_N - \beta_0} \right| \overset{a.s.}{\Rightarrow} 0, \quad (11)$$

where the a.s. uniform convergence result follows from Jennrich’s (1969) uniform strong law of large numbers. Moreover,

$$\left| E \left[ \exp \left( -t. \left( X' \xi + ||X|| + W - V - \alpha_0 \right) \right) \right]_{\xi = \hat{\beta}_N - \beta_0} \right| \overset{a.s.}{\Rightarrow} 0.$$
\[-E \left[ \exp \left( -t \left( ||X|| + W - V - \alpha_0 \right) \right) \right] \]
\[\leq E \left[ \left| \exp \left( -t \left( X'\xi + ||X|| \right) \right) - \exp(-t||X||) \right|_{\xi=\hat{\beta}_N - \beta_0} \times E \left[ \exp \left( -t \left( W - V - \alpha_0 \right) \right) \right] \]
\[\leq E \left[ \left| \exp \left( -t \left. X'\xi \right| - 1 \cdot \exp(-t||X||) \right|_{\xi=\hat{\beta}_N - \beta_0} \times E \left[ \exp \left( -t \left( W - V - \alpha_0 \right) \right) \right] \right] \]
\[\leq ||\hat{\beta}_N - \beta_0||t.E[||X||]\exp(t.\alpha_0)\exp(\sigma_0^2t^2/2)E[\exp(-t.W)] , \]
where the last inequality follows from the mean value theorem. It follows now straightforwardly from (6) that

\[
\sup_{t \in [0,c]} \left| E \left[ \exp \left( -t \left( X'\xi + ||X|| + W - V - \alpha_0 \right) \right) \right|_{\xi=\hat{\beta}_N - \beta_0} - E[\exp(-t||X||)] \exp(t.\alpha_0)\exp(\sigma_0^2t^2/2)E[\exp(-t.W)] \right| \overset{a.s.}{\to} 0. \tag{12}
\]

Since \( \frac{1}{N} \sum_{i=1}^{N} \exp(-t||X_i||) \overset{a.s.}{\to} E[\exp(-t||X||)] \) uniformly on \([0,c]\) as well, it follows now from (11) and (12) that the following key result hold.

**Lemma 1.** Denote

\[Z_{i,N} = X_{i} \hat{\beta}_N - Y_{i} . \tag{13}\]

Under Assumption 1, and for the linear SF model (1),

\[
\sup_{t \in [0,c]} \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t \left( Z_{i,N} + ||X_i|| \right) \right) \right| \left. - \exp(t.\alpha_0)\exp(\sigma_0^2t^2/2).E[\exp(-t.W)].E[\exp(-t||X||)] \right| \overset{a.s.}{\to} 0 \tag{14}
\]

for any constant \( c > 0. \)

Since \( \frac{1}{N} \sum_{i=1}^{N} \exp(-t||X_i||) \overset{a.s.}{\to} E[\exp(-t||X||)] \) uniformly on \([0,c]\), it follows now trivially from Lemma 1 that the following result holds. C.f. (3).
Theorem 1. Denote

\[ \Upsilon_N(t) = \ln \left( \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t(Z_{i,N} + ||X_i||) \right) \right) - \ln \left( \frac{1}{N} \sum_{i=1}^{N} \exp (-t.||X_i||) \right). \]

Then under Assumption 1, for any constant \( c > 0 \),

\[ \sup_{t \in [0,c]} \left| \Upsilon_N(t) - \alpha_0 t - \sigma_0^2 t^2/2 - \ln \left( E[\exp(-t.W)] \right) \right| \xrightarrow{a.s.} 0 \]

and thus also

\[ \int_0^c (\Upsilon_N(t) - \alpha_0 t - \sigma_0^2 t^2/2 - \ln \left( E[\exp(-t.W)] \right))^2 \, dt \xrightarrow{a.s.} 0. \]

Note that the linearity of model (1) is convenient but not essential for these results. If the model is nonlinear, say

\[ Y = \alpha_0 + g(X, \beta_0) + V - W \]
\[ = \mu_0 + g(X, \beta_0) + V - (W - E[W]) \]
\[ \mu_0 = \alpha_0 - E[W], \]

where \( g(x, \beta) \) is a known function (for example the log of a CES production function), then under Assumption 1 and standard non-linear least squares (NLLS) conditions we can estimate \( \mu_0 \) and \( \beta_0 \) strongly consistent and \( \sqrt{N} \) asymptotic normally by NLLS. C.f. Jennrich (1969). Then with \( \hat{\beta}_N \) the NLLS estimator of \( \beta_0 \) and under standard NLLS conditions it is not hard to verify that the result (14) carries over for \( Z_{i,N} = g(X_i, \hat{\beta}_N) - Y_i \) and \( ||X_i|| \) replaced by \( \sup_{\beta \in B} ||\partial g(X_i, \beta)/\partial \beta|| \), where \( B \) is a given compact and convex parameter space containing \( \beta_0 \) in its interior. With these modifications, all the results in this paper carry over to nonlinear SF models.

4 Identification: The general case

Suppose that there exists an \( \alpha \), a \( \sigma^2 > 0 \) and a nonnegative random variable \( W_* \) such that

\[ \sup_{t \in [0,c]} \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t. (Z_{i,N} + ||X_i||) \right) \right| - \exp(t.\alpha + \sigma^2 t^2/2).E[\exp(-t.W_*).E[\exp(-t.||X||)]] \xrightarrow{a.s.} 0 \]
as well. C.f. Lemma 1, and multiplying the absolute value by the weight function \( \eta(t) = t^2 \exp(-t^3/3) \), for example, the result involved holds for \( c = \infty \) as well:

\[
\sup_{t > 0} \eta(t) \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t.(Z_{i,N} + ||X_i||) \right) \right. \\
- \exp(t.\alpha + \sigma^2 t^2 / 2).E[\exp(-t.W_*)].E[\exp(-t.||X||)] \xrightarrow{a.s.} 0 \quad (17)
\]

and similar by Lemma 1,

\[
\sup_{t > 0} \eta(t) \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t.(Z_{i,N} + ||X_i||) \right) \right. \\
- \exp(t.\alpha_0 + \sigma^2_0 t^2 / 2).E[\exp(-t.W_*)].E[\exp(-t.||X||)] \xrightarrow{a.s.} 0
\]

Since the Laplace transform \( E[\exp(-t.||X||)] \) can be estimated strongly consistent uniformly on \([0, \infty)\) we may without loss of generality assume, for the purpose of identification, that \( E[\exp(-t.||X||)] \) is known.

Now (17) implies that for all \( t > 0 \),

\[
\exp(t.\alpha_0 + \sigma^2_0 t^2 / 2).E[\exp(-t.W)] \\
\equiv \exp(t.\alpha + \sigma^2 t^2 / 2).E[\exp(-t.W_*)], \quad (18)
\]

or equivalently,

\[
E[\exp(-t.W)] \\
\equiv \exp(t.(\alpha - \alpha_0) + (\sigma^2 - \sigma^2_0)t^2/2).E[\exp(-t.W_*)], \quad (19)
\]

\[
E[\exp(-t.W_*)] \\
\equiv \exp(t.(\alpha_0 - \alpha) + (\sigma^2_0 - \sigma^2)t^2/2)E[\exp(-t.W)]. \quad (20)
\]

To prove that (18) implies \( \sigma^2 = \sigma^2_0 \), suppose in first instance that \( \sigma^2 > \sigma^2_0 \). Then by (19)

\[
E[\exp(-t.W)] \\
\geq \exp \left( t.(\alpha - \alpha_0) + (\sigma^2 - \sigma^2_0)t^2/2 \right).E[\exp(-t.W_*)].I(W_* \leq K) \\
\geq \exp \left( t.(\alpha - \alpha_0 - K) + (\sigma^2 - \sigma^2_0)t^2/2 \right).\Pr [W_* \leq K],
\]
where $I(.)$ is the well-known indicator function and $K$ is chosen such that $\Pr[ W_\ast \leq K ] > 0$. Letting $t \to \infty$ the right-hand side of this inequality converges to $\infty$, whereas the left-hand side converges to 0, so that in the limit this inequality would read $0 \geq \infty$. Consequently, $\sigma^2 > \sigma_0^2$ is not possible. By a similar argument, (20) implies that $\sigma^2 < \sigma_0^2$ is not possible either, so that $\sigma^2 = \sigma_0^2$.

Thus, (19) and (20) become
\begin{align}
E[\exp(-t.W)] &\equiv \exp(t.(\alpha - \alpha_0)).E[\exp(-t.W_\ast)], \\
E[\exp(-t.W_\ast)] &\equiv \exp(t.(\alpha_0 - \alpha))E[\exp(-t.W)], 
\end{align}
respectively.

Suppose first that $\alpha_0 > \alpha$. Then it follows from (22) that
\begin{align}
E[\exp(-t.W_\ast)] &\geq \exp(t.(\alpha_0 - \alpha))E[\exp(-t.W)I(W \leq (\alpha_0 - \alpha)/2] \\
&\geq \exp(t.(\alpha_0 - \alpha)/2) \Pr[W \leq (\alpha_0 - \alpha)/2],
\end{align}
where by Assumption 1(e), $\Pr[W \leq (\alpha_0 - \alpha)/2] > 0$. Again, letting $t \to \infty$ leads to the contradiction $0 \geq \infty$, so that $\alpha_0 - \alpha > 0$ is not possible.

Now suppose that $\alpha_0 < \alpha$. Then by (22), similar to (23), we have the inequality
\begin{align}
E[\exp(-t.W)] &\geq \exp(t.(\alpha - \alpha_0)/2). \Pr[W_\ast \leq (\alpha - \alpha_0)/2],
\end{align}
so that by the same argument as before, $\alpha_0 < \alpha$ is not possible, provided that $\Pr[W_\ast \leq (\alpha - \alpha_0)/2] > 0$. However, the latter condition is not guaranteed. Thus, at this point we can only conclude that $\gamma = \alpha - \alpha_0 \geq 0$, so that (22) reads
\begin{align}
E[\exp(-t.W_\ast)] &\equiv E[\exp(-t.(W + \gamma))], \forall t > 0.
\end{align}

As is well-known, two nonnegative random variables have the same distribution if and only if their Laplace transforms are equal. Thus, (24) implies that $W_\ast \sim W + \gamma$. This case corresponds to the following version of SF model (1):
\begin{align}
Y = \alpha + X'/\beta_0 + V - W_\ast = \alpha_0 + \gamma + X'/\beta_0 + V - (W + \gamma).
\end{align}
Of course, $\gamma$ cancels out in (25), so that the latter model is observational equivalent to (1). Only if we can impose (or assume) the condition
\begin{align}
\Pr[W_\ast \leq w] > 0 \text{ for all } w > 0
\end{align}
can we conclude that $\alpha = \alpha_0$, and then by the equality $E[\exp(-t.W_\ast)] = E[\exp(-t.W)]$ for all $t > 0$ it follows that $W_\ast \sim W$.

Summarizing, the following identification result has been shown.

**Theorem 2.** Let $\mathcal{F}_+^t$ be the collection of all distribution functions $F$ of nonnegative random variables satisfying $F(w) > 0$ for $w > 0$, and let $F_0(w)$ be the distribution function of the inefficiency variable $W$ in SF model (1). Suppose that the latter model is observational equivalent to the alternative model $Y = \alpha + X'\beta_0 + V_\ast - W_\ast$, where $X$, $V_\ast$ and $W_\ast$ are independent, $V_\ast \sim \mathcal{N}(0, \sigma^2)$, and the c.d.f. $F$ of $W_\ast$ belongs to $\mathcal{F}_+$. Then under Assumption 1, $\alpha = \alpha_0$, $\sigma^2 = \sigma^2_0$ and $F = F_0$.

Note that apart from the condition $F(w) > 0$ for $w > 0$ there are no further restrictions on $F \in \mathcal{F}_+$, so that Theorem 2 applies to nonnegative discrete, continuous and mixed discrete-continuous distributions for $W$.

In principle we could use the expression between the absolute value bars in (17) as the basis for an objective function, by taking the square and then integrating it relative to an appropriate weight function, for example $\eta(t) = t^2 \exp(-t^3/3)$. A practical problem with this idea is that in this case it is too difficult to concentrate $\alpha$ and $\sigma^2$ out. This is the very reason why we propose to base our SNP estimation approach on the objective function (4). However, the proof of Theorem 2 is based on letting $t \to \infty$ whereas in (4) $t$ is confined to $[0, c]$. Therefore, in the latter case the identification proof has to be redone from scratch.

5 SNP modeling the densities on $[0, \infty)$ and their Laplace transforms

From now onwards we will only focus on the following case.

**Assumption 2.** Let $\mathcal{F}_{a+}$ be the collection of all absolutely continuous distribution functions $F(w)$ with density $f(w)$ and support $S = \{w \geq 0 : f(w) > 0\} \subset [0, \infty)$. The true c.d.f. $F_0(w)$ of $W$ is a member of $\mathcal{F}_{a+}$ but its density $f_0(w)$ is continuous and positive on $(0, \infty)$.

---

8Possibly except singular distributions.
Thus, each c.d.f. \( F \in \mathcal{F}_{\alpha^+} \) takes the form \( F(w) = \int_0^w f(x)dx \), but the density \( f(w) \) may be discontinuous, for example if \( F(w) \) is piecewise linear, and/or \( f(w) \) may be equal to zero on a subset of \([0, \infty)\) with positive but finite Lebesgue measure. An example of the latter case if \( F(w) = F_0(\gamma + w) \) for some positive constant \( \gamma \), for which \( f(w) = 0 \) for \( w < \gamma \). C.f. the case (25), and that its support \( S \) is such that \([0, \infty) \setminus S \) has finite Lebesgue measure.9

5.1 Hilbert spaces of functions

Given a density \( \omega(x) \) with support \( S(\omega) = \{ x \in \mathbb{R} : \omega(x) > 0 \} \), consider the space \( L^2(\omega) \) of Borel measurable real functions \( f(x) \) on \( S(\omega) \) satisfying \( \int f(x)^2 \omega(x)dx < \infty \). Endow \( L^2(\omega) \) with the inner product \( \langle f, g \rangle = \int f(x)g(x)\omega(x)dx \) and associated norm \( ||f|| = \sqrt{\langle f, f \rangle} \) and metric \( ||f - g|| \), for \( f, g \in L^2(\omega) \). Under mild conditions on \( \omega \) it can be shown that every Cauchy sequence in \( L^2(\omega) \) takes a limit in \( L^2(\omega) \), which makes \( L^2(\omega) \) a Hilbert space, by definition. Thus, a Hilbert space mimics the properties of a Euclidean space, which is therefore also a Hilbert space.

Moreover, under the same mild conditions it can be shown that \( L^2(\omega) \) is separable, which means the there exits an orthonormal sequence \( \{ \rho_m \}_{m=0}^\infty \) in \( L^2(\omega) \), i.e., \( \langle \rho_m, \rho_k \rangle = I(m = k) \), with \( I(.) \) the indicator function, such that for every \( f \in L^2(\omega) \), and with \( f_n \) defined as \( f_n(x) = \sum_{m=0}^n \langle \rho_m, f \rangle \cdot \rho_m(x) \), we have \( \lim_{n \to \infty} ||f - f_n|| = 0 \). Such an orthonormal sequence \( \{ \rho_m \}_{m=0}^\infty \) is called complete in \( L^2(\omega) \).

In particular, if \( \int |x|^m \omega(x)dx < \infty \) for all \( m \in \mathbb{N} \) then \( \omega \) generates a unique sequence (up to sign) \( \{ \rho_m \}_{m=0}^\infty \) of orthonormal polynomials of order \( m \), respectively, that is complete in \( L^2(\omega) \). These polynomials obey the three-term recurrence relation (TTRR)

\[
a_{k+1} \rho_{k+1}(x) + (b_k - x) \rho_k(x) + a_k \cdot \rho_{k-1}(x) = 0, \quad k \in \mathbb{N},
\]

starting from \( \rho_0(x) = 1 \) and \( \rho_1(x) = (x - c_1)/\sqrt{c_2} \), where \( c_1 = \int x \omega(x)dx \) and \( c_2 = \int (x - c_1)^2 \omega(x)dx \), and the sequences \( a_k \) and \( b_k \) are specific for \( \omega(x) \). See for example Hamming (1973) for the TTRR (27), and Bierens (2014a, Theorem 11) for the completeness \( \{ \rho_m \}_{m=0}^\infty \).

In the case that \( \omega(x) \) is the standard normal density the orthonormal polynomials \( \{ \rho_m \}_{m=0}^\infty \) generated by \( \omega(x) \) are known as the Hermite poly-

---

9The latter condition excludes singular distributions.
nomials, for which $\rho_0(x) = 1$, $\rho_1(x) = x$, and $a_k = \sqrt{k}$, $b_k = 0$ in TTRR (27).

The standard exponential density, $\omega(x) = \exp(-x)$, $x \geq 0$, generates the orthonormal Laguerre polynomials $\{\rho_m\}_{m=0}^{\infty}$, for which $\rho_0(x) = 1$, $\rho_1(x) = x - 1$, and $a_k = k$, $b_k = 2k + 1$ in TTRR (27).

The uniform $[0,1]$ density, i.e. $\omega(x) = I(0 < x < 1)$, generates the orthonormal Legendre polynomials $\{\rho_m\}_{m=0}^{\infty}$, for which $\rho_0(x) = 1$, $\rho_1(x) = \sqrt{3}(2x - 1)$, and $a_k = 0.5k/\sqrt{4k^2 - 1}$, $b_k = 0.5$ in (27).

In the latter uniform case the Hilbert space $L^2(\omega)$ is usually denoted by $L^2(0,1)$. Another, non-polynomial, complete orthonormal sequences in $L^2(0,1)$ is the cosine sequence: $\rho_0(x) = 1$, $\rho_k(x) = \sqrt{2}\cos(k\pi x)$ for $k \in \mathbb{N}$ and $x \in [0,1]$.

A few more examples of complete orthonormal sequences are listed in Bierens (2014a), and many more in other books and papers on Hilbert spaces.

5.2 SNP modeling of density functions

Next, let $f(x)$ be a density with support $S(\omega)$ or smaller. At this point we will not impose any further conditions on $f$ other than the required Borel measurability. Then $g = \sqrt{f/\omega} \in L^2(\omega)$, and given a complete orthonormal sequence $\{\rho_m\}_{m=0}^{\infty}$ in $L^2(\omega)$ we can approximate $g(x)$ by $g_n(x) = \sum_{m=0}^{n} \langle \rho_m, g \rangle \rho_m(x)$, in the sense that

$$
||g - g_n||^2 = \int \left( \sqrt{f(x)/\omega(x)} - \sum_{m=0}^{n} \langle \rho_m, \sqrt{f/\omega} \rangle \cdot \rho_m(x) \right)^2 \omega(x)dx
$$

$$
= \int \left( \sqrt{f(x)} - \sqrt{\omega(x)} \sum_{m=0}^{n} \langle \rho_m, \sqrt{f/\omega} \rangle \cdot \rho_m(x) \right)^2 dx
$$

$$
= \int (\sqrt{f(x)} - h_n(x))^2 dx \to 0 \text{ as } n \to \infty
$$

where $h_n(x) = \sqrt{\omega(x)}\sum_{m=0}^{n} \langle \rho_m, \sqrt{f/\omega} \rangle \cdot \rho_m(x)$. This suggests to approximate $f(x)$ by

$$
f_n^*(x) = h_n(x)^2 = \omega(x) \left( \sum_{m=0}^{n} \kappa_m \cdot \rho_m(x) \right)^2
$$
where \( \kappa_m = \langle \rho_m, \sqrt{f}/\sqrt{\omega} \rangle = \int \rho_m(x) \sqrt{\omega(x)} \sqrt{f(x)} \, dx \). Note that \( \kappa_m^2 < 1 \). However, \( \int f_n^*(x) \, dx = \sum_{m=0}^{\infty} \kappa_m^2 < 1 \) because by (28), \( \sum_{m=0}^{\infty} \kappa_m^2 = \int f(x) \, dx = 1 \), as is not hard to verify. This can be corrected by dividing \( f_n^*(x) \) by \( \sum_{m=0}^{\infty} \kappa_m^2 \), so that \( f_n(x) = \omega(x) (\sum_{m=0}^{\infty} \kappa_m \rho_m(x))^2 (\sum_{k=0}^{\infty} \kappa_k^2)^{-1} \) is a proper density function.

In most cases the first element \( \rho_0 \) of the sequence \( \{\rho_m\}_{m=0}^{\infty} \) is identical to 1. Then \( \kappa_0 = \int \sqrt{\omega(x)} \sqrt{f(x)} \, dx \in (0, 1) \) hence, denoting \( \delta_m = \kappa_m/\kappa_0 \) for \( m \in \mathbb{N} \), \( f_n(x) \) reads

\[
f_n(x) = \omega(x) \frac{1 + \sum_{m=1}^{\infty} \delta_m \rho_m(x)}{1 + \sum_{k=1}^{\infty} \delta_k^2}.
\] (29)

where \( \sum_{k=1}^{\infty} \delta_k^2 = \kappa_0^{-1} - 1 \in (0, \infty) \). Moreover, it is not hard to verify that

\[
\lim_{n \to 0} \int |f(x) - f_n(x)| \, dx = 0.
\] (30)

This result implies, by Theorems 9 and 11 in Bierens (2014a), that \( \lim_{n \to \infty} f_n(x) = f(x) \) a.e.\(^{11} \) on the support \( S \) of \( \omega(x) \), so that

\[
f(x) = \omega(x) \frac{1 + \sum_{m=1}^{\infty} \delta_m \rho_m(x)}{1 + \sum_{k=1}^{\infty} \delta_k^2} \text{ a.e., } \sum_{k=1}^{\infty} \delta_k^2 < \infty.
\] (31)

However, it follows from Bierens (2014a, Theorem 16) that in general the sequence \( \{\delta_m\}_{m=1}^{\infty} \) is not unique: there exist possibly uncountable many sequences \( \{\delta_m\}_{m=1}^{\infty} \) for which (30) and (31) hold. On the other hand, if the support \( S(\omega) \) of \( f(x) \) is connected and if \( f(x) \) is continuous and positive on the interior of \( S(\omega) \) then \( \{\delta_m\}_{m=1}^{\infty} \) is unique. See Bierens (2014a, Theorem 21). Recall that by Assumption 2 these conditions apply to the true density \( f_0 \) of the inefficiency variable \( W \).

In their seminal paper, Gallant and Nychka (1987) used a bivariate version\(^{12} \) of (29) on the basis of Hermite polynomials, to approximate the bivariate error density of the latent variable equations in Heckman’s (1979) sample

\( \sqrt{\int \varphi_m^2(x) \omega(x) \sqrt{f(x)} \, dx} \left| \int \varphi_m(x) \sqrt{\omega(x)} \sqrt{f(x)} \, dx \right| < \sqrt{\int \varphi_m^2(x) \omega(x) \, dx} \sqrt{\int f(x) \, dx} = ||\varphi_m|| = 1 \)

\( ^{10} \)Because by Schwarz inequality \( \left| \int \varphi_m(x) \sqrt{\omega(x)} \sqrt{f(x)} \, dx \right| < \sqrt{\int \varphi_m^2(x) \omega(x) \, dx} \sqrt{\int f(x) \, dx} = ||\varphi_m|| = 1 \)

\( ^{11} \)The abbreviation “a.e.” stands for “almost everywhere”, which means that the limit result involved holds pointwise, except perhaps on a set with zero Lebesgue measure.

\( ^{12} \)See Bierens (2014a, Theorem 17).
selection model in order to estimate this model semi-nonparametrically. Since then the univariate version of (29), with \( \omega(x) \) the standard normal density and the \( \rho_m(x) \)'s the Hermite polynomials, has become the standard approach in the econometric SNP literature, except the following.

In the SNP models considered in Bierens (2008) and Bierens and Carvalho (2007) the unknown function involved is a distribution function on the unit interval, which they model semi-nonparametrically by numerically integrating the corresponding SNP density (29) with \( \omega(x) = 1 \) and the \( \rho_m(x) \)'s the Legendre polynomials. The auction model considered in Bierens and Song (2012) involves the unknown absolutely continuous c.d.f. \( F(v) \) with density \( f(v) \) on \([0, \infty)\) of the individual values, which they convert to a distribution function on the unit interval by using the following trick. For an a priori chosen absolutely continuous distribution function \( G(v) \) with support \([0, \infty)\) or \((0, \infty)\), density \( g(v) \) and inverse \( G^{-1}(u) \), one can write \( F(v) = H(G(v)) \), where \( H(u) = F(G^{-1}(u)) \) is now an unknown absolutely continuous c.d.f. on \([0, 1]\) with density \( h(u) = f(G^{-1}(u))/g(G^{-1}(u)) \). Again, this density is modeled as (29) with \( \omega(x) = 1 \) and the \( \rho_m(x) \)'s the Legendre polynomials. Bierens (2014b) used the same trick, with \( G \) now a c.d.f. with support \( \mathbb{R} \), in particular the logistic c.d.f., but instead of Legendre polynomials for SNP modeling of \( h(u) \) he used the cosine sequence, because then the corresponding SNP version of \( H(u) \) has a closed form expression, and so do its derivatives.

5.3 SNP densities based on Laguerre polynomials and their parameter space

In the present paper the unknown density \( f_0(w) \) of the inefficiency variable \( W \) is the main object of interest, which appears indirectly as its Laplace transform. Therefore, the SNP density (29) with \( x = w \), \( \omega(w) = \exp(-w) \) and the \( \rho_m(w) \)'s the Laguerre polynomials seems the best option for SNP modeling of \( f_0(w) \) and any competing density \( f(w) \). C.f. Assumption 2.

In this case we can write the right-hand side of (31) as

\[
f(w|\delta) = \exp(-w)\frac{1 + \sum_{m=1}^{\infty} \delta_m \rho_m(w))^2}{1 + \sum_{k=1}^{\infty} \delta_k^2}, \quad w \geq 0,
\]

where \( \delta = \{\delta_m\}_{m=1}^{\infty} \) is an infinite-dimensional parameter contained in the
parameter space

\[ \Delta = \left\{ \pmb{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}, \tag{33} \]

and the \( \rho_m(w) \)'s are the Laguerre polynomials. Recall that the Laguerre polynomials can be generated recursively by

\[ \rho_{m+1}(w) = \frac{1}{m+1} (w - 1) \rho_m(w) - \frac{m}{m+1} \left( 2 \rho_m(w) + \rho_{m-1}(w) \right), \]

\( m \in \mathbb{N} \), starting from \( \rho_0(w) = 1, \rho_1(w) = w - 1. \) (34)

Moreover, the Laguerre polynomials have the closed form expression

\[ \rho_m(w) = \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{(-1)^\ell}{\ell!} w^\ell. \tag{35} \]

The parameter space \( \Delta \) will be endowed with the innerproduct

\[ \langle \pmb{\delta}_1, \pmb{\delta}_2 \rangle = \sum_{m=1}^{\infty} \delta_{1,m} \delta_{2,m}, \quad \pmb{\delta}_i = \{\delta_{i,m}\}_{m=1}^{\infty} \text{ for } i = 1, 2, \tag{36} \]

and associated pseudo-Euclidean norm

\[ ||\pmb{\delta}|| = \sqrt{\langle \pmb{\delta}, \pmb{\delta} \rangle} = \sqrt{\sum_{m=1}^{\infty} \delta_m^2} \tag{37} \]

and metric \( ||\pmb{\delta}_1 - \pmb{\delta}_2|| \), so that \( \Delta \) becomes a Hilbert space.\(^{13}\)

Next, denote by \( \pi_n \) the truncation operator, i.e.,

**Definition 1.** \( \pi_n \) applied to \( \pmb{\delta} = \{\delta_m\}_{m=1}^{\infty} \) as \( \pi_n \pmb{\delta} \) replaces all the \( \delta_m \)'s with \( m > n \) by zeros.

Then by (32), for \( n \in \mathbb{N}, \)

\[ f(w|\pi_n \pmb{\delta}) = \exp(-w) \frac{(1 + \sum_{m=1}^{n} \delta_m \rho_m(w))^2}{1 + \sum_{k=1}^{n} \delta_k^2}, \quad w \geq 0, \tag{38} \]

\(^{13}\)The latter follows from the fact that every Cauchy sequence in \( \Delta \) converges to a point in \( \Delta \), which is not hard to verify.
so that by (30), $\lim_{n \to \infty} \int_0^\infty |f(w|\delta) - f(w|\pi_n \delta)|dw = 0$. More precisely, we have

$$\int_0^\infty |f(w|\pi_n \delta) - f(w|\delta)|dw \leq 2||\delta - \pi_n \delta||^2 + 4||\delta - \pi_n \delta|| \to 0 \quad (39)$$

as $n \to \infty$, where the limit follows from the fact that $\sum_{m=1}^\infty \delta^2_m < \infty$ implies $||\delta - \pi_n \delta||^2 = \sum_{m=n+1}^\infty \delta^2_m \to 0$, and the inequality is due to the following more general lemma.

**Lemma 2.** For every pair $\delta_1, \delta_2$ in $\Delta$, $\int_0^\infty |f(w|\delta_1) - f(w|\delta_2)|dw \leq 2||\delta_1 - \delta_2||^2 + 4||\delta_1 - \delta_2||$.

**Proof.** Appendix.

### 5.4 Laplace transforms

The result (31) now reads: Given any density $f(w)$ of an absolutely continuous distribution on $[0, \infty)$, there exist possibly uncountable many parameters $\delta \in \Delta$ such that $f(w) = f(w|\delta)$ a.e. on $[0, \infty)$.\(^{14}\) On the other hand, if $f(w)$ is continuous and positive on $(0, \infty)$ then there exists a unique $\delta \in \Delta$ such that $f(w) = f(w|\delta)$ a.e. on $[0, \infty)$.\(^ {15}\) The latter implies, by Assumption 2, that for the true density $f_0(w)$ of the inefficiency variable $W$ there exists a unique $\delta^0 \in \Delta$ such that $f_0(w) = f(w|\delta^0)$ a.e. on $[0, \infty)$. Since the caveat ”a.s.” no longer applies after integration, it follows that in general the Laplace identity

$$\int_0^\infty \exp(-tw)f(w)dw \equiv \int_0^\infty \exp(-tw)f(w|\delta)dw, \ t > 0,$$

holds for at least one $\delta \in \Delta$, but possibly for uncountable many $\delta$’s in $\Delta$ as well, whereas the Laplace identity

$$\int_0^\infty \exp(-tw)f_0(w)dw \equiv \int_0^\infty \exp(-tw)f(w|\delta^0)dw, \ t > 0$$

only holds only for a single $\delta^0 \in \Delta$.


\(^{15}\)C.f. Bierens (2014a, Theorem 21).
Finally, it follows trivially from (39) that
\[
\sup_{t>0} \left| \int_0^\infty \exp(-t.w) f(w|\delta)dw - \int_0^\infty \exp(-t.w) f(w|\pi_n\delta)dw \right|
\leq 2||\delta - \pi_n\delta||^2 + 4||\delta - \pi_n\delta|| \to 0
\]
as \(n \to \infty\).

6 The SNP objective function and its asymptotic properties

As alluded to by equation (4) in the introduction, our estimation approach is based on the empirical objective function
\[
Q_N(\alpha, \sigma^2, \delta|c) = \int_0^c \left( \Upsilon_N(t) - \alpha t - \sigma^2 t^2/2 - \ln(L(t|\delta)) \right)^2 dt
\]
\[
= \int_0^c \left( \Psi_N(t|\delta) - \alpha t - \sigma^2 t^2/2 \right)^2 dt, \quad (40)
\]
for some given constant \(c > 0\), where now \(\Upsilon_N(t)\) is defined in Theorem 1,
\[
L(t|\delta) = \int_0^\infty \exp(-t.w) f(w|\delta)dw \quad (41)
\]
is the Laplace transform of \(f(w|\delta)\), and
\[
\Psi_N(t|\delta) = \Upsilon_N(t) - \ln(L(t|\delta)). \quad (42)
\]

Note that by Theorem 1, pointwise in \(\alpha, \sigma^2, \delta\),
\[
Q_N(\alpha, \sigma^2, \delta|c) \xrightarrow{a.s.} Q(\alpha, \sigma^2, \delta|c), \text{ where}
\]
\[
Q(\alpha, \sigma^2, \delta|c) = \int_0^c \left( \Lambda(t|\delta) - (\alpha - \alpha_0)t - (\sigma^2 - \sigma_0^2)t^2/2 \right)^2 dt, \quad (43)
\]
with
\[
\Lambda(t|\delta) = \ln(L(t|\delta^0)) - \ln(L(t|\delta)). \quad (44)
\]
The next step is to concentrate \(\alpha\) and \(\sigma^2\) out:
Lemma 3. Given \( \delta \in \Delta \), the solutions of the minimization problem

\[
(\alpha_N(\delta|c), \sigma^2_N(\delta|c)) = \arg \min_{(\alpha, \sigma^2)} Q_N(\alpha, \sigma^2, \delta|c)
\]

are

\[
\alpha_N(\delta|c) = 48c^{-3}\int_0^c \tau \Psi_N(\tau|\delta) d\tau - 60c^{-4}\int_0^c \tau^2 \Psi_N(\tau|\delta) d\tau,
\]

\[
\sigma^2_N(\delta|c) = -120c^{-4}\int_0^c \tau \Psi_N(\tau|\delta) d\tau + 160c^{-5}\int_0^c \tau^2 \Psi_N(\tau|\delta) d\tau,
\]

where \( \Psi_N(t|\delta) \) is defined by (42). Moreover, under the conditions of Theorem 1 it follows that \( \alpha_N(\delta|c) \xrightarrow{a.s.} \alpha(\delta|c) \) and \( \sigma^2_N(\delta|c) \xrightarrow{a.s.} \sigma^2(\delta|c) \), pointwise in \( \delta \in \Delta \), where

\[
\alpha(\delta|c) = \alpha_0 + 48c^{-3}\int_0^c \tau \Lambda(\tau|\delta) d\tau - 60c^{-4}\int_0^c \tau^2 \Lambda(\tau|\delta) d\tau,
\]

\[
\sigma^2(\delta|c) = \sigma_0^2 - 120c^{-4}\int_0^c \tau \Lambda(\tau|\delta) d\tau + 160c^{-5}\int_0^c \tau^2 \Lambda(\tau|\delta) d\tau,
\]

hence, \( \alpha(\delta^0|c) = \alpha_0 \) and \( \sigma^2(\delta^0|c) = \sigma_0^2 \). Furthermore, \( \alpha(\delta|c) \) and \( \sigma^2(\delta|c) \) are just the solutions of

\[
(\alpha(\delta|c), \sigma^2(\delta|c)) = \arg \min_{(\alpha, \sigma^2)} Q(\alpha, \sigma^2, \delta|c).
\]

The solutions (45) and (46) follow straightforwardly from the first-order conditions for \( Q_N(\alpha, \sigma^2, \delta|c) \) and similar for (47) and (48). The a.s. convergence results follow from Theorem 1.

As said before, the actual objective function for the estimation of \( \delta^0 \) is

\[
Q_N(\delta|c) = Q_N(\alpha_N(\delta|c), \sigma^2_N(\delta|c), \delta|c)
\]

\[
= \int_0^c \left( \Psi_N(t|\delta) - \alpha_N(\delta|c)t - \sigma^2_N(\delta|c)t^2/2 \right)^2 dt,
\]

for which it follows straightforwardly from Theorem 1 and Lemma 3 that pointwise in \( \delta \in \Delta \),

\[
Q_N(\delta|c) \xrightarrow{a.s.} Q(\delta|c) = \inf_{\alpha, \sigma^2} Q(\alpha, \sigma^2, \delta|c), \text{ where}
\]

\[
Q(\delta|c) = \int_0^c \left( \Lambda(t|\delta) - (\alpha(\delta|c) - \alpha_0)t - (\sigma^2(\delta|c) - \sigma_0^2)t^2/2 \right)^2 dt.
\]
Moreover, it follows trivially that \( Q(\delta^0|c) = 0 \). But is \( \delta^0 \) is unique, i.e., does \( Q(\delta|c) = 0 \) imply \( \delta = \delta^0 \)? The answer is Yes, as will be shown in the next section.

7 Identification: The absolute continuous case

To show that \( \delta^0 \) is unique, we need to show that the set

\[
\Delta_0(c) = \{ \delta \in \Delta : Q(\delta|c) = 0 = 0 \}
\]

equals the singleton \( \{ \delta^0 \} \).

Note that by (50), \( \delta \in \Delta_0(c) \) implies that for all \( t \in (0, c) \),

\[
0 = t.(\alpha(\delta|c) - \alpha_0) + \frac{1}{2}t^2(\sigma^2(\delta|c) - \sigma^2_0) + \Lambda(t|\delta).
\]

Taking the derivative to \( t \in (0, c) \) and then letting \( t \downarrow 0 \) yield

\[
\alpha(\delta|c) - \alpha_0 = -\Lambda'(t|\delta)|_{t=0}
\]

\[
= \int_0^\infty w \exp(-t.w) f(w|\delta)dw - \int_0^\infty w \exp(-t.w) f(w|\delta^0)dw \bigg|_{t=0} - \int_0^\infty w \exp(-t.w) f(w|\delta^0)dw
\]

\[
= \int_0^\infty w f(w|\delta)dw - \int_0^\infty w f(w|\delta^0)dw. \quad (51)
\]

Similarly,

\[
\sigma^2(\delta|c) - \sigma^2_0 = -\Lambda''(t|\delta)
\]

\[
= -\frac{\int_0^\infty w^2 \exp(-t.w) f(w|\delta)dw}{\int_0^\infty \exp(-t.w) f(w|\delta)dw} + \left( \frac{\int_0^\infty w \exp(-t.w) f(w|\delta)dw}{\int_0^\infty \exp(-t.w) f(w|\delta)dw} \right)^2
\]

\[
+ \frac{\int_0^\infty w^2 \exp(-t.w) f(w|\delta^0)dw}{\int_0^\infty \exp(-t.w) f(w|\delta^0)dw} - \left( \frac{\int_0^\infty w \exp(-t.w) f(w|\delta^0)dw}{\int_0^\infty \exp(-t.w) f(w|\delta^0)dw} \right)^2, \quad (52)
\]

hence, letting \( t \downarrow 0 \),

\[
\sigma^2(\delta|c) - \sigma^2_0 = \int_0^\infty w^2 f(w|\delta^0)dw - \left( \int_0^\infty w f(w|\delta^0)dw \right)^2
\]

\[
- \int_0^\infty w^2 f(w|\delta)dw + \left( \int_0^\infty w f(w|\delta)dw \right)^2. \quad (53)
\]
Thus, given \( \delta \in \Delta_0(c) \), \( a(\delta) = \alpha(\delta|c) - \alpha_0 \) and \( b(\delta) = \sigma^2(\delta|c) - \sigma_0^2 \) are constants which do not depend on \( c \), so that

\[
\Lambda(t|\delta) = -a(\delta).t - \frac{1}{2}b(\delta).t^2
\]

(54)

\[
= \ln \left( \int_0^\infty \exp(-t.w)f(w|\delta)dw \right) - \ln \left( \int_0^\infty \exp(-t.w)f(w|\delta^0)dw \right),
\]

hence, by (52) and (54),

\[
a(\delta) = -\Lambda'(t|\delta) + t.\Lambda''(t|\delta)
\]

(55)

for all \( t \in (0, c) \).

7.1 **Proof of** \( \alpha(\delta|c) = \alpha_0 \)

We will now show that \( \delta \in \Delta_0(c) \) implies \( a(\delta) = 0 \), as follows. Substituting (52) and (55) in (54) now yields \( \Lambda(t|\delta) = t.\Lambda'(t|\delta) - \frac{1}{2}t^2\Lambda''(t|\delta) \), which is a second-order homogeneous differential equation:

\[
\Lambda''(t|\delta) = 2t^{-1}.\Lambda'(t|\delta) - 2t^{-2}\Lambda(t|\delta).
\]

(56)

As is well-known, the general solution of (56) is of the form \( \Lambda(t|\delta) = p.t^q \), with \( \Lambda'(t|\delta) = p.q.t^{q-1} \) and \( \Lambda''(t|\delta) = p.q(q-1).t^{q-2} \), so that \( p.q(q-1).t^{q-2} = 2p.q.t^{q-2} - 2p.t^{q-2} \). If \( p \neq 0 \) the latter equation implies that either \( q = 1 \) or \( q = 2 \). But by (52), \( \Lambda''(t|\delta) = p.q(q-1).t^{q-2} = -b(\delta) \), so that \( q = 2 \) and \( p = -b(\delta)/2 \). Thus, the solution of (56) is

\[
\Lambda(t|\delta) = -b(\delta).t^2/2
\]

(57)

which by (54) implies that \( a(\delta) = 0 \). It follows now from (51) that

\[
\delta \in \Delta_0(c) \text{ implies } \int_0^\infty w.f(w|\delta)dw = \int_0^\infty w.f(w|\delta^0)dw.
\]

(58)
7.2 Proof of $\sigma^2(\delta|c) = \sigma_0^2$

Next, we will show that $\delta \in \Delta_0(c)$ implies $b(\delta) = 0$ as well. Observe first from (53) and (58) that

$$b(\delta) = \int_0^\infty w^2 f(w|\delta^0)dw - \int_0^\infty w^2 f(w|\delta)dw.$$  \hspace{1cm} (59)

Moreover, observe from (44) and (57) that

$$\int_0^\infty \exp(-t.w)f(w|\delta^0)dw = \exp\left(b(\delta).t^2/2\right)\int_0^\infty \exp(-t.w)f(w|\delta)dw$$  \hspace{1cm} (60)

Given a $t \in (0,c)$, let $x > 0$ be such that $t + x < c$. Then

$$\int_0^\infty \exp(-x.w)\exp(-t.w)f(w|\delta^0)dw = \exp\left(b(\delta).(t + x)^2/2\right)\int_0^\infty \exp(-x.w)\exp(-t.w)f(w|\delta)dw$$

$$= (\exp\left(b(\delta).x(x + 2t)/2\right) - 1)\exp\left(b(\delta).t^2/2\right)\int_0^\infty \exp(-x.w)\exp(-t.w)f(w|\delta)dw$$

as well. Since $\exp(-x.w) = 1 + \sum_{m=1}^\infty (-1)^m x^m w^m/m!$ and similarly

$$\exp\left(b(\delta).x(x + 2t)/2\right) - 1 = \sum_{m=1}^\infty \frac{1}{m!}\left(b(\delta)/2\right)^m x^m (x + 2t)^m$$

it follows from this equality, together with (60), that

$$\sum_{m=1}^\infty \frac{(-1)^m x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta^0)dw$$

$$- \exp\left(b(\delta).t^2/2\right) \sum_{m=1}^\infty \frac{(-1)^m x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta)dw$$

$$= \exp\left(b(\delta).t^2/2\right) \sum_{k=1}^\infty \frac{1}{k!}\left(b(\delta)/2\right)^k x^k (x + 2t)^k$$
\[
\times \int_0^\infty (\exp(-x.w) - 1) \exp(-t.w)f(w|\delta)dw \\
+ \exp \left( b(\delta) t^2 / 2 \right) \sum_{k=1}^\infty \frac{1}{k!} (b(\delta)/2)^k x^k(x + 2t)^k \int_0^\infty \exp(-t.w)f(w|\delta)dw \\
= \exp \left( b(\delta) t^2 / 2 \right) \sum_{k=1}^\infty \frac{1}{k!} (b(\delta)/2)^k x^k(x + 2t)^k \\
\times \sum_{m=1}^\infty (-1)^m \frac{x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta)dw \\
+ \exp \left( b(\delta) t^2 / 2 \right) \sum_{k=1}^\infty \frac{1}{k!} (b(\delta)/2)^k x^k(x + 2t)^k \\
\times \sum_{m=1}^\infty (-1)^m \frac{x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta)dw \\
\}

Dividing both sides of this equation by \( x \) and then letting \( x \to 0 \) yield

\[
\int_0^\infty w \exp(-t.w)f(w|\delta^0)dw + \exp \left( b(\delta) t^2 / 2 \right) \int_0^\infty w \exp(-t.w)f(w|\delta)dw \\
= t.b(\delta) \exp \left( b(\delta) t^2 / 2 \right) \int_0^\infty \exp(-t.w)f(w|\delta)dw \\
\]

hence

\[
\sum_{m=2}^\infty (-1)^m \frac{x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta^0)dw \\
- \exp \left( b(\delta) t^2 / 2 \right) \sum_{m=2}^\infty (-1)^m \frac{x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta)dw \\
= \exp \left( b(\delta) t^2 / 2 \right) \sum_{k=1}^\infty \frac{1}{k!} (b(\delta)/2)^k x^k(x + 2t)^k \\
\times \sum_{m=1}^\infty (-1)^m \frac{x^m}{m!} \int_0^\infty w^m \exp(-t.w)f(w|\delta)dw \\
+ \exp \left( b(\delta) t^2 / 2 \right) \sum_{k=1}^\infty \frac{1}{k!} (b(\delta)/2)^k x^k(x + 2t)^k \int_0^\infty \exp(-t.w)f(w|\delta)dw \\
+ x^2 \exp \left( b(\delta) t^2 / 2 \right) (b(\delta)/2) \int_0^\infty \exp(-t.w)f(w|\delta)dw \\
\]

Next, dividing both sides of this equation by \( x^2 / 2 \) and then letting \( x \to 0 \)}
yields
\[\int_{0}^{\infty} w^2 \exp(-t.w) f(w|\delta^0)dw \]
\[- \exp (b(\delta).t^2/2) \int_{0}^{\infty} w^2 \exp(-t.w) f(w|\delta)dw \]
\[= \exp (b(\delta).t^2/2) b(\delta) \int_{0}^{\infty} w \exp(-t.w) f(w|\delta)dw \]
\[+ \exp (b(\delta).t^2/2) (b(\delta)/2)^2 (2t)^2 \int_{0}^{\infty} \exp(-t.w) f(w|\delta)dw \]
\[+ \exp (b(\delta).t^2/2) b(\delta) \int_{0}^{\infty} \exp(-t.w) f(w|\delta)dw.\]

Finally, letting \( t \to 0 \) in this equation yields
\[\int_{0}^{\infty} w^2 f(w|\delta^0)dw - \int_{0}^{\infty} w^2 f(w|\delta)dw = b(\delta) \left( \int_{0}^{\infty} w f(w|\delta)dw + 1 \right),\]
which by (59) and (58) implies that \( b(\delta) \int_{0}^{\infty} w f(w|\delta^0)dw = 0 \). Thus,
\( \delta \in \Delta_0(c) \) implies \( b(\delta) = 0 \).

### 7.3 Uniqueness of \( \delta^0 \)

It follows now from (54) that for \( \delta \in \Delta_0(c) \),
\[\int_{0}^{\infty} \exp(-t.w) f(w|\delta^0)dw \equiv \int_{0}^{\infty} \exp(-t.w) f(w|\delta)dw \text{ on } (0,c).\]

The left side integral is the Laplace transform of the actual distribution of \( W \), \( E[\exp(-t.W)] = \int_{0}^{\infty} \exp(-t.w) f(w|\delta^0)dw \), and the right side integral is the Laplace transform of an alternative nonnegative distribution, represented by \( W^* \), say, \( E[\exp(-t.W^*)] = \int_{0}^{\infty} \exp(-t.w) f(w|\delta^0)dw \), so that
\[E[\exp(-t.W)] \equiv E[\exp(-t.W^*)] \text{ on } (0,c). \quad (61)\]

As is well-known, if two Laplace transforms are equal on an arbitrary open
subset of \((0, \infty)\) then the corresponding distributions are equal.\(^{16}\) Thus, (61) implies \(W \sim W_*\), hence \(\int_0^w f(x|\delta^0)dx = \int_0^w f(x|\delta)dx\) for all \(w > 0\), which in its turn implies that
\[
f(w|\delta^0) = f(w|\delta) \text{ for all } w > 0.
\] (62)

Recall that for the true density \(f_0(w)\) of the inefficiency variable \(W\) there exists a unique \(\delta^0 \in \Delta\) such that \(f_0(w) = f(w|\delta^0)\) a.e. on \([0, \infty)\), so that (62) implies \(\delta = \delta^0\). Consequently, \(\Delta_0(c) = \{\delta^0\}\).

Summarizing the following identification result has been proved.

**Theorem 3.** Under Assumptions 1 and 2 the infinite-dimensional parameter \(\delta^0 \in \Delta\) in the SNP representation \(f_0(w) = f(w|\delta^0)\) a.e. on \([0, \infty)\) of the true density \(f_0(w)\) of the inefficiency variable \(W\) is identified as the only \(\delta \in \Delta\) for which \(Q_N(\delta|c) \overset{a.s.}{\to} 0\). Thus, \(f_0(w)\) itself is semi-nonparametrically identified.

Note that the identification problem in the case (25) is no longer a problem, as the current approach automatically forces \(\gamma = 0\) in the latter model.

8 **Integrated method of moments sieve estimation**

8.1 **Sieve estimation**

First note that the number of solutions of \(\arg \min_{\delta \in \Delta} Q_N(\delta)\) is infinite because \(\Delta\) is infinite dimensional, and that neither of these solutions will be consistent. The standard approach to avoid these problems is sieve estimation, as follows.

The idea of sieve estimation, proposed by Grenander (1981), is to construct an increasing sequence of finite dimensional compact subsets \(\Delta_n\),

\(^{16}\)If not well-known, here is a quick proof. For \(t\) in an open subset of \((0, \infty)\) we can infinitely many times take the derivatives of (61), resulting in the equalities \(E[W^m \exp(-t.W)] = E[W_*^m \exp(-t.W_*)]\) for \(m \in \mathbb{N}\). Then for arbitrary \(\xi \in \mathbb{R}\) and with \(i = \sqrt{-1}\) it follows from the series expansion of the complex \(\exp(\cdot)\) function that \(E[\exp(i\xi W) \exp(-t.W)] = \sum_{m=0}^{\infty} (1/m!)(i\xi)^m E[W^m \exp(-t.W)] = E[\exp(i\xi W_*) \exp(-t.W_*)]\). Letting \(t \downarrow 0\) yields \(E[\exp(i\xi W)] = E[\exp(i\xi W_*)]\), which implies \(W \sim W_*\) by the uniqueness of characteristic functions.
\( n \in \mathbb{N}, \) of \( \Delta, \) called sieve spaces, such that \( \bigcup_{n=1}^{\infty} \Delta_n = \Delta, \) where here the bar denotes the closure. For example, let

\[
\Delta_n = \left\{ \delta = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{n} \delta_m^2 \leq M_n, \ \delta_m = 0 \text{ for } m > n \right\},
\]

where \( M_n \) is a given strictly increasing positive sequence converging to \( \infty \) as \( n \to \infty. \) Next, let \( n_N \in \mathbb{N} \) be any nondecreasing subsequence of the sample size \( N \) satisfying

\[
\lim_{N \to \infty} n_N = \infty, \ \lim_{N \to \infty} n_N / N = 0. \tag{63}
\]

Then

\[
\hat{\delta}_N = \arg \min_{\delta \in \Delta_N} Q_N(\delta) \tag{64}
\]

is a sieve estimator of \( \delta^0, \) for which it will be shown, in the next subsection, that the following consistency results hold.

**Theorem 4.** Under Assumptions 1 and 2 and the condition (63), the sieve estimator is \( \hat{\delta}_N \) weakly consistent, in the sense that \( \|\hat{\delta}_N - \delta^0\| \overset{p}{\to} 0 \) as \( N \to \infty. \) This result implies that \( \hat{\alpha}_N = \alpha_N(\hat{\delta}_N) \overset{p}{\to} \alpha_0, \ \hat{\sigma}_N^2 = \sigma_N^2(\hat{\delta}_N) \overset{p}{\to} \sigma_0^2, \) and by Lemma 2,

\[
\int_0^{\infty} |f(w|\hat{\delta}_N) - f_0(w)|dw \leq 2\|\delta^0 - \hat{\delta}_N\|^2 + 4\|\delta^0 - \hat{\delta}_N\| \overset{p}{\to} 0.
\]

**8.2 Consistency of the sieve estimators**

The consistency of sieve estimators is well-established in the literature. See for example Chen (2007) for a review, and Bierens (2014b, Theorem 4.2) for an alternative approach. However, these results require rather high-level conditions. On the other hand, in the present case the consistency results in Theorem 4 can be proved much easier than in Chen (2007) and Bierens (2014b).

In general a crucial condition for the consistency of sieve estimators is that the objective function, \( Q_N(\delta) \) in our case, and its pointwise limit \( Q(\delta) \) are continuous in the parameter \( \delta. \) However, in the present case we only need the continuity of \( Q(\delta) \) in \( \delta^0, \) which follows straightforwardly from (47), (48), (50) and Lemma 2.

Next, consider the following easy inequalities.
Lemma 4. For any $(\alpha, \sigma^2, \delta) \in \mathbb{R} \times [0, \infty) \times \Delta$,
\[
Q_N(\alpha, \sigma^2, \delta|c) \leq \sqrt{Q(\alpha, \sigma^2, \delta|c)} 
\left(\sqrt{Q(\alpha, \sigma^2, \delta|c)} + R_N\right),
\]
\[
\geq \sqrt{Q(\alpha, \sigma^2, \delta|c)} \left(\sqrt{Q(\alpha, \sigma^2, \delta|c)} - R_N\right),
\]
with $R_N = 2\sqrt{Q_N(\alpha_0, \sigma^2_0, \delta^0|c)}$,
for $Q_N(\alpha, \sigma^2, \delta|c)$ and $Q(\alpha, \sigma^2, \delta|c)$ defined by (40) and (43), respectively.

Proof. See the Appendix. ■

Note that by Theorem 1, $R_N \overset{a.s.}{\rightarrow} 0$. Now denote for an arbitrary $\varepsilon > 0$,
\[
\Xi(\varepsilon) = \{\delta \in \Delta : ||\delta - \delta^0|| \geq \varepsilon\}. \tag{65}
\]
Then by the second inequality in Lemma 4, $\hat{\delta}_N \in \Xi(\varepsilon)$ implies
\[
\underline{Q}_N(\hat{\delta}_N) = Q_N(\alpha_N(\hat{\delta}_N), \sigma_N^2(\hat{\delta}_N), \hat{\delta}_N|c) 
\geq \sqrt{Q(\alpha_N(\hat{\delta}_N), \sigma_N^2(\hat{\delta}_N), \hat{\delta}_N|c)} 
\times \left(\sqrt{Q(\alpha_N(\hat{\delta}_N), \sigma_N^2(\hat{\delta}_N), \hat{\delta}_N|c)} - R_N\right)
\]
Moreover, $\hat{\delta}_N \in \Xi(\varepsilon)$ implies that
\[
Q(\alpha_N(\hat{\delta}_N), \sigma_N^2(\hat{\delta}_N), \hat{\delta}_N|c) \geq Q(\alpha(\hat{\delta}_N), \sigma^2(\hat{\delta}_N), \hat{\delta}_N|c) 
= \underline{Q}(\hat{\delta}_N) \geq \inf_{\delta \in \Xi(\varepsilon)} \underline{Q}(\delta) = \kappa,
\]
say, where $\kappa > 0$, whereas by $R_N \overset{a.s.}{\rightarrow} 0$ there exists an $N_0$ such that $\sqrt{\kappa}/2 > R_N$ a.s. for all $N \geq N_0$. Thus, for $N \geq N_0$, $\hat{\delta}_N \in \Xi(\varepsilon)$ implies $Q_N(\hat{\delta}_N) \geq \kappa/2$, hence
\[
\Pr[\hat{\delta}_N \in \Xi(\varepsilon)] \leq \Pr[Q_N(\hat{\delta}_N) \geq \kappa/2]. \tag{66}
\]
On the other hand, with $N$ so large that $\pi_n N \delta^0 \in \Delta_{nN}$, which is the case for $M_{nN} > ||\delta^0||^2$, we have by the first inequality in Lemma 4, that
\[
Q_N(\hat{\delta}_N) \leq Q_N(\pi_n N \delta^0)
\]
28
\[
Q_N(\alpha_N(\pi_n, \delta^0), \sigma^2_N(\pi_n, \delta^0), \pi_n, \delta^0 | c) \leq Q_N(\alpha(\pi_n, \delta^0), \sigma^2(\pi_n, \delta^0), \pi_n, \delta^0 | c)
\]

\[
\leq \sqrt{Q(\alpha(\pi_n, \delta^0), \sigma^2(\pi_n, \delta^0), \pi_n, \delta^0 | c)} \times \left( \sqrt{Q(\pi_n, \delta^0 | c)} + R_N \right)
\]

\[
= \sqrt{Q(\pi_n, \delta^0 | c)} \left( \sqrt{Q(\pi_n, \delta^0 | c)} + R_N \right)
\]

\[
\overset{\text{a.s.}}{\to} Q(\delta^0 | c) = 0
\]

by continuity of \( Q(\delta | c) \) in \( \delta^0 \) and \( R_N \overset{\text{a.s.}}{\to} 0 \). Consequently, \( \lim_{N \to \infty} \Pr[Q_N(\delta_N) \geq \kappa/2] = 0 \), so that by (66), \( \lim_{N \to \infty} \Pr[\delta_N \in \Xi(\varepsilon)] = 0 \), or equivalently, \( \lim_{N \to \infty} \Pr[||\delta_N - \delta^0|| \geq \varepsilon] = 0 \). Since \( \varepsilon > 0 \) was arbitrary, the main result \( \lim_{N \to \infty} \Pr[||\delta_N - \delta^0|| \overset{p}{\to} 0] \) follows. The results \( \alpha_N(\delta_N) \overset{p}{\to} \alpha_0 \) and \( \sigma^2_N(\delta_N) \overset{p}{\to} \sigma^2_0 \) follow from Theorem 1 and Lemma 2, as is not hard to verify.

\section{Prediction of inefficiency}

For each firm \( i \in N \) we have, under Assumption 1,

\[
Y_i - \tilde{\beta}_N X_i - \tilde{\alpha}_N = \alpha_0 - \tilde{\alpha}_N - (\tilde{\beta}_N - \beta_0)' X_i + V_i - W_i
\]

\[
= V_i - W_i + o_p(1),
\]

as is easy to verify. Thus, for each \( i \in N \), \( U_i = V_i - W_i \) can be estimated consistently by \( \hat{U}_{i,i} = Y_i - \tilde{\beta}_N X_i - \tilde{\alpha}_N \).

Recall that \( \exp(-W_i) \) is the relative distance of firm \( i \) from its production frontier. Since \( U_i = V_i - W_i \) may be treated as being known, the conditional expectation \( E[\exp(-W_i)|U_i] \), known as the technical efficiency (TE) index for firm \( i \),\(^{17}\) is the best estimate of \( \exp(-W_i) \), ”best” in the sense that the conditional mean square error \( E[(\exp(-W_i) - E[\exp(-W_i)|U_i])^2|U_i] \) is minimal.

As to the computation of the TE index under Assumptions 1-2, recall that the joint density of \((V,W)\) takes the form

\[
f_{V,W}(v, w) = \frac{1}{\sigma_0} \phi(v/\sigma_0) f_0(w),
\]

\(^{17}\)See for example Jondrow et al. (1982) and Horrace and Parmeter (2018).
where $\phi$ is the standard normal density, hence the joint density of $W$ and $U = V - W$ takes the form

$$f_{W,U}(w,u) = \frac{1}{\sigma_0} \phi \left( \frac{u + w}{\sigma_0} \right) f_0(w),$$

with marginal densities $f_W(w) = \int_{-\infty}^{\infty} f_{W,U}(w,u)du = f_0(w)$ and

$$f_U(u) = \int_{0}^{\infty} f_{W,U}(w,u)dw = \int_{0}^{\infty} \frac{1}{\sigma_0} \phi \left( \frac{u + w}{\sigma_0} \right) f_0(w)dw. \tag{68}$$

Hence,

$$E[\exp(-W_i)|U_i = u] = \int_{0}^{\infty} \exp(-w) f_{W,U}(w,u)dw = \frac{f_U(u)}{f_U(u)} = \gamma_0(u), \text{ say.}$$

Clearly, the TE function $\gamma_0(u)$ is continuous and uniformly bounded on $\mathbb{R}$.

Replacing $\sigma_0$ by its estimate $\widehat{\sigma}_N$ and $f_0(w)$ by its estimate $f(w|\widehat{\delta}_N)$ yield

$$\widehat{\gamma}_N(u) = \frac{\int_{0}^{\infty} \exp(-w) f_{N,W,U}(w,u)dw}{f_{N,U}(u)}$$

as an estimate of the TE function $\gamma_0(u)$, where

$$\widehat{f}_{N,W,U}(w,u) = \frac{1}{\widehat{\sigma}_N} \phi \left( \frac{u + w}{\widehat{\sigma}_N} \right) f(w|\widehat{\delta}_N),$$

$$\widehat{f}_{N,U}(u) = \int_{0}^{\infty} \frac{1}{\widehat{\sigma}_N} \phi \left( \frac{u + w}{\widehat{\sigma}_N} \right) f(w|\widehat{\delta}_N)dw,$$

for which the following results hold.

**Theorem 5.** Under Assumptions 1-2, $\sup_{|u| \leq M} |\widehat{\gamma}_N(u) - \gamma_0(u)| = o_p(1)$ for any $M \in (0, \infty)$ and $\widehat{\gamma}_N(\widehat{U}_{N,i}) - \gamma_0(U_i) = o_p(1)$ for each $i \in \mathbb{N}$.

**Proof.** See the Appendix. ■

Note that $\gamma_0(U_i) = E[\exp(-W_i)|U_i] = E[\exp(-W_i)|V_i - W_i]$ is the actual TE index of firm $i$, so that the last result in Theorem 5 reads

$$\widehat{\gamma}_N(Y_i - \widehat{\beta}_N X_i - \widehat{\alpha}_N) = E[\exp(-W_i)|V_i - W_i] + o_p(1).$$
10 A numerical experiment

To be done!

11 An empirical application

In this section, we apply our SNP approach to estimate the wage frontier for men and the impact of human capital on the industrial wage distributions in Taiwan. In the relevant studies by Polachek and Yoon (1987), Hofler and Murphy (1992), Polachek and Robst (1998), the stochastic frontier model is used to measure a worker’s incomplete information about available wages based on job searching theory, where the incomplete information is defined as the difference between a worker’s observed wage and his/her maximum potential wage. In other words, the observed wage rates fall below the maximum potential wage offers because of costly job search.

We use the data taken from the 2013 Taiwan’s Manpower Utilization Survey (Directorate-General of Budget, Accounting and Statistics), which includes 12,252 men in the labor market. The sample is divided into three industries (groups): Industry 1 (service industry), Industry 2 (manufacturing industry), and Industry 3 (other industries) that contain, respectively, 3176, 4574, and 4502 observations in each industry. The dependent variable is the logarithm of monthly wage (Y), and the independent variables include the human capital variables: years of base schooling (Edu1), years of higher education (Edu2), experience (Exp), and the squared experience (Exp2). A dummy variable for the firm scale (Scale) is also included as the firm size effect on wage. More detail about the variable definitions are referred to Table 1 below.

The model is specified as

\[ Y^j = \alpha^j_0 + \beta^j_1 \text{Edu}_1 + \beta^j_2 \text{Edu}_2 + \beta^j_3 \text{Exp} + \beta^j_4 \text{Exp}^2 + \beta^j_5 \text{Scale} + V^j - W^j \]

where \( j = 1, 2, 3 \) denotes the industry. The parameters involved, including the variance of \( V^j \) and the distribution of \( W^j \), are allowed to be different across industries, so the model will be estimated for each industry separately.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage</td>
<td>The monthly wage of male worker in year 2013</td>
</tr>
<tr>
<td>Edu₁</td>
<td>Years of base-education schooling (including elementary, junior and senior high schools)</td>
</tr>
<tr>
<td>Edu₂</td>
<td>Years of higher-education schooling (including college, graduate school)</td>
</tr>
<tr>
<td>Exp</td>
<td>Years of experience, which is defined as ( \text{Exp} = \min{\text{Age} - 15, \text{Age} - \text{Years of Education} - 8} )</td>
</tr>
<tr>
<td>Exp²</td>
<td>Squared experience divided by 100, i.e., ( (\text{Exp} \times \text{Exp})/100 )</td>
</tr>
<tr>
<td>Scale</td>
<td>A dummy variable. Scale = 1, if the number of the employees is larger than 100; and zero otherwise.</td>
</tr>
</tbody>
</table>

To be completed!
12 Appendix

12.1 Proof of Lemma 2

Observe from (32) that for every pair $\delta_1 = \{\delta_{1,m}\}_{m=1}^{\infty}, \delta_2 = \{\delta_{2,m}\}_{m=1}^{\infty}$ in $\Delta$,

\[
f(w|\delta_1) - f(w|\delta_2) \\
= \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right) \\
- \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) \\
= \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right) - \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) \\
+ \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right) - \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) \\
= \left( 1 - \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\delta) \\
+ \frac{1}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
\times \left( 2 + \sum_{m=1}^{\infty} (\delta_{1,m} + \delta_{2,m}) \rho_m(w) \right) \\
= \left( \frac{\sum_{m=1}^{\infty} (\delta_{2,m}^2 - \delta_{1,m}^2)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\delta) \\
+ \frac{1}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
\times \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) + 2 \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w) \right) \\
+ \frac{2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
= \left( \frac{\sum_{m=1}^{\infty} (\delta_{2,m}^2 - \delta_{1,m}^2)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\delta)
\]
Next, observe that
\[
\sum_{m=1}^\infty (\delta_{2,m}^2 - \delta_{1,m}^2) = \sum_{m=1}^\infty (\delta_{2,m}^2 - \delta_{1,m}) (\delta_{2,m} + \delta_{1,m}) \\
\leq \sum_{m=1}^\infty (\delta_{2,m}^2 - \delta_{1,m})^2 + 2 \sum_{m=1}^\infty (\delta_{2,m} - \delta_{1,m}) \delta_{2,m}^2 \\
\leq ||\delta_2 - \delta||^2 + 2 ||\delta_2 - \delta|| ||\delta_2||
\]
where the second equality follows from (36) and (37), and the last inequality follows from the well-known Cauchy-Schwarz inequality.

It follows now, using the Cauchy-Schwarz and Lyapunov’s inequalities and the orthonormality of the Laguerre sequence \(\rho_m(w)\) with respect to \(\exp(-w)\) that
\[
\int_0^\infty |f(w|\delta_1) - f(w|\delta_2)| \, dw \\
\leq \frac{||\delta_2 - \delta_1||^2 + 2 ||\delta_2 - \delta|| ||\delta_2||}{1 + ||\delta_2||^2} \\
+ \frac{1}{1 + ||\delta_2||^2} \int_0^{\infty} \exp(-w) \left( \sum_{m=1}^\infty (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right)^2 \, dw \\
+ \frac{2}{1 + ||\delta_2||^2} \int_0^{\infty} \exp(-w) \left| \sum_{m=1}^\infty (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right| \, dw
\]
\[
\left. \times \left| \sum_{m=1}^{\infty} \delta_{2,m}^2 \rho_m(w) \right| \, dw \rightvert
\]
\[
\frac{2}{1 + \| \delta_2 \|^2} \int_0^\infty \exp(-w) \left| \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right| \, dw
\]
\[
\leq \frac{1}{1 + \| \delta_2 \|^2} \left( \| \delta_2 - \delta_1 \|^2 + 2 \| \delta_2 - \delta_1 \| \| \delta_2 \| \right)
\]
\[
+ \int_0^\infty \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right)^2 \, dw
\]
\[
+ 2 \left( \int_0^\infty \exp(-w) \left( \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w) \right)^2 \, dw \right)^{\frac{1}{2}} \left( \int_0^\infty \exp(-w) \left( \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w) \right)^2 \, dw \right)^{\frac{1}{2}}
\]
\[
\times \left. \int_0^\infty \exp(-w) \left( \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w) \right)^2 \, dw \right| \left. \int_0^\infty \exp(-w) \left( \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w) \right)^2 \, dw \right| \right.
\]
\[
\leq 2 \| \delta_2 - \delta_1 \|^2 + 4 \| \delta_2 - \delta_1 \| \| \delta_2 \| + 2 \| \delta_2 - \delta_1 \|
\]
\[
\leq 2 \| \delta_2 - \delta_1 \|^2 + 4 \| \delta_2 - \delta_1 \| \| \delta_2 \| + 2 \| \delta_2 - \delta_1 \|
\]
\[
\leq 2 \| \delta_2 - \delta_1 \|^2 + 4 \| \delta_2 - \delta_1 \|
\]

where the last inequality follows from the fact that \( \frac{d}{dx}(1 + x^2)^{-1} = (1 - x^2)/(1 + x^2)^2 = 0 \) for \( x = 1 \), hence
\[
\frac{\| \delta_2 \|}{1 + \| \delta_2 \|^2} \leq \max_{x \geq 0} x(1 + x^2)^{-1} = \frac{1}{2},
\]

35
12.2 Proof of Lemma 4

Recall from (40) and (43) that $Q_N(\alpha, \sigma^2, \delta|c) = \int_0^c \psi_N(t|\alpha, \sigma^2, \delta)^2 dt$, where

$$
\psi_N(t|\alpha, \sigma^2, \delta) = \gamma_N(t) - \alpha t - \sigma^2 t^2/2 - \ln(L(t|\delta)),
$$

and $Q(\alpha, \sigma^2, \delta|c) = \int_0^c \psi(t|\alpha, \sigma^2, \delta)^2 dt$, where

$$
\psi(t|\alpha, \sigma^2, \delta) = \ln(L(t|\delta^0)) - \ln(L(t|\delta)) - (\alpha - \alpha_0)t - (\sigma^2 - \sigma_0^2) t^2/2.
$$

Then it follows trivially that

$$
Q_N(\alpha, \sigma^2, \delta|c) - Q(\alpha, \sigma^2, \delta|c) = \int_0^c \left( \psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta) \right) \times \left( \psi_N(t|\alpha, \sigma^2, \delta) + \psi(t|\alpha, \sigma^2, \delta) \right) dt
$$

$$
= \int_0^c \left( \psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta) \right)^2 dt + 2 \int_0^c \left( \psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta) \right) \psi(t|\alpha, \sigma^2, \delta) dt
$$

$$
= Q_N(\alpha_0, \sigma_0^2, \delta^0|c) + 2 \int_0^c \psi_N(t|\alpha_0, \sigma_0^2, \delta^0) \psi(t|\alpha, \sigma^2, \delta) dt,
$$

because trivially, $\psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta) = \psi_N(t|\alpha_0, \sigma_0^2, \delta^0)$. Since

$$
\left| \int_0^c \psi_N(t|\alpha_0, \sigma_0^2, \delta^0) \psi(t|\alpha, \sigma^2, \delta) dt \right| \leq \sqrt{\int_0^c \psi_N(t|\alpha_0, \sigma_0^2, \delta^0)^2 dt} \sqrt{\int_0^c \psi(t|\alpha, \sigma^2, \delta)^2 dt}
$$

$$
= \sqrt{Q_N(\alpha_0, \sigma_0^2, \delta^0|c)} \sqrt{Q(\alpha, \sigma^2, \delta|c)}
$$

the inequalities in Lemma 4 follow.

12.3 Proof of Theorem 5

It will first be shown that

$$
\sup_{u \in \mathbb{R}} \left| \widetilde{f}_{N,U}(u) - f_U(u) \right| = o_p(1), \quad (69)
$$

36
as follows. Observe that

\[
\left| \hat{f}_{N,U}(u) - f_U(u) \right| \\
= \left| \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{u+w}{\sigma_N} \right) f(w)\tilde{\delta}_N dw - \int_0^\infty \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) f_0(w) dw \right| \\
\leq \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{u+w}{\sigma_N} \right) \left| f(w)\tilde{\delta}_N - f_0(w) \right| dw \\
+ \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{u+w}{\sigma_N} \right) \left| \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) - \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) \right| f_0(w) dw \\
\leq \frac{1}{\sigma_N \sqrt{2\pi}} \int_0^\infty \left| f(w)\tilde{\delta}_N - f_0(w) \right| dw \\
+ \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{u+w}{\sigma_N} \right) \left| \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) - \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) \right| f_0(w) dw. \quad (70)
\]

If \( \tilde{\sigma}_N < \sigma_0 \) then

\[
\left| \frac{1}{\sigma_N} \phi \left( \frac{u+w}{\sigma_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{u+w}{\sigma_0} \right) \right| \\
\leq \exp \left( -\frac{1}{2}u^2/\tilde{\sigma}_N^2 \right) - \exp \left( -\frac{1}{2}u^2/\sigma_0^2 \right) \\
+ \frac{\exp \left( -\frac{1}{2}u^2/\sigma_0^2 \right)}{\tilde{\sigma}_N \sqrt{2\pi}} - \frac{\exp \left( -\frac{1}{2}u^2/\sigma_0^2 \right)}{\sigma_0 \sqrt{2\pi}} \\
\leq \exp \left( -\frac{1}{2}(u+w)^2(\tilde{\sigma}_N^2 - \sigma_0^2) \right) - 1 \left| \frac{\exp \left( -\frac{1}{2}(u+w)^2/\sigma_0^2 \right)}{\tilde{\sigma}_N \sqrt{2\pi}} \right| \\
+ \frac{\tilde{\sigma}_N^{-1} - \sigma_0^{-1}}{\sqrt{2\pi}} \exp \left( -\frac{1}{2}(u+w)^2/\sigma_0^2 \right) \\
+ \frac{1}{\sqrt{2\pi}} \left| \tilde{\sigma}_N^{-1} - \sigma_0^{-1} \right| \\
\leq \frac{|\tilde{\sigma}_N^{-2} - \sigma_0^{-2}|\sigma_0^2}{\tilde{\sigma}_N \sqrt{2\pi}} \cdot \frac{(u+w)^2}{2\sigma_0^2} \exp \left( -\frac{1}{2}(u+w)^2/\sigma_0^2 \right) + \frac{1}{\sqrt{2\pi}} \left| \tilde{\sigma}_N^{-1} - \sigma_0^{-1} \right| \\
\leq \frac{|\tilde{\sigma}_N^{-2} - \sigma_0^{-2}|\sigma_0^2}{\tilde{\sigma}_N \sqrt{2\pi}} \exp(-1) + \frac{1}{\sqrt{2\pi}} \left| \tilde{\sigma}_N^{-1} - \sigma_0^{-1} \right|, \quad (71)
\]
where the fourth inequality follows from the mean value theorem, and the last inequality follows from \( \sup_{x>0} x \exp(-x) = \exp(-1) \).

Similarly, if \( \hat{\sigma}_N > \sigma_0 \) then
\[
\begin{align*}
\left| \frac{1}{\hat{\sigma}_N} \phi \left( \frac{u + w}{\hat{\sigma}_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{u + w}{\sigma_0} \right) \right| & \leq \exp \left( -\frac{1}{2} \frac{(u + w)^2 / \hat{\sigma}_N^2}{\hat{\sigma}_N \sqrt{2\pi}} \right) - \frac{\exp \left( -\frac{1}{2} \frac{(u + w)^2 / \hat{\sigma}_N^2}{\sigma_0 \sqrt{2\pi}} \right)}{\sigma_0 \sqrt{2\pi}} \\
& \leq \frac{\exp \left( -\frac{1}{2} \frac{(u + w)^2 / \hat{\sigma}_N^2}{\sigma_0 \sqrt{2\pi}} \right)}{\sigma_0 \sqrt{2\pi}} \left[ 1 - \exp \left( -\frac{1}{2} (u + w)^2 (\sigma_0^{-2} - \hat{\sigma}_N^{-2}) \right) \right] \\
& + \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right| \exp \left( -\frac{1}{2} \frac{(u + w)^2 / \hat{\sigma}_N^2}{\sigma_0 \sqrt{2\pi}} \right) \\
& \leq \frac{\left| \hat{\sigma}_N^{-2} - \sigma_0^{-2} \right| \sigma_0^{-2} (u + w)^2}{\sigma_0 \sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{(u + w)^2 / \hat{\sigma}_N^2}{\sigma_0 \sqrt{2\pi}} \right) + \frac{1}{\sqrt{2\pi}} \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right| \\
& \leq \frac{\left| \hat{\sigma}_N^{-2} - \sigma_0^{-2} \right| \sigma_0^{-2} (u + w)^2}{\sigma_0 \sqrt{2\pi}} \exp(-1) + \frac{1}{\sqrt{2\pi}} \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right|. \tag{72}
\end{align*}
\]
It follows now straightforwardly from (70), (71), (72) and Theorem 3 that (69) holds.

Moreover, similar to (69) it follows that
\[
\sup_{u \in \mathbb{R}} \left| \hat{f}_{N,U}(u) \hat{\gamma}_N(u) - f_U(u) \gamma_0(u) \right| = \sup_{u \in \mathbb{R}} \left| \int_0^\infty \exp(-w) \hat{f}_{N,W,U}(w,u)dw \right. \left. - \int_0^\infty \exp(-w) f_{W,U}(w,u)dw \right| = o_p(1),
\]
hence,
\[
\begin{align*}
f_U(u) \left| \hat{\gamma}_N(u) - \gamma_0(u) \right| & = \left| \hat{f}_{N,U}(u) \hat{\gamma}_N(u) - f_U(u) \gamma_0(u) - \left( \hat{f}_{N,U}(u) - f_U(u) \right) \hat{\gamma}_N(u) \right| \\
& \leq \sup_{u \in \mathbb{R}} \left| \hat{f}_{N,U}(u) \hat{\gamma}_N(u) - f_U(u) \gamma_0(u) \right| + \sup_{u \in \mathbb{R}} \left| \hat{f}_{N,U}(u) - f_U(u) \right| \\
& = o_p(1), \tag{73}
\end{align*}
\]
where the inequality is due to \( \sup_{u \in \mathbb{R}} \gamma_N(u) \leq 1 \).

Now observe from (68) that by bounded convergence, \( \lim_{|u| \to \infty} f_U(u) = 0 \), but \( f_U(u) > 0 \) for all \( u \in \mathbb{R} \). Since \( f_U(u) \) is continuous on \( \mathbb{R} \) it follows therefore that for any constant \( M > 0 \), \( \inf_{|u| \leq M} f_U(u) > 0 \), which in its turn implies by (73) that

\[
\sup_{|u| \leq M} |\gamma_N(u) - \gamma_0(u)| \leq \frac{\sup_{|u| \leq M} f_U(u) \cdot |\gamma_N(u) - \gamma_0(u)|}{\inf_{|u| \leq M} f_U(u)} = o_p(1). \tag{74}
\]

As has been shown before, for each \( i \), \( U_i = V_i - W_i \) can be estimated consistently by \( \widehat{U}_{N,i} = Y_i - \beta_{N+1} X_i - \alpha_N \). It will now be shown that this implies that \( \gamma_N(\widehat{U}_{N,i}) - \gamma_0(U_i) = o_p(1) \), as follows. Let \( \varepsilon \in (0,1) \) be arbitrary, and let \( M > 0 \) be so large that \( \Pr[|U_i| > M/2] < \varepsilon/2 \), which implies \( \Pr[|U_i| > M] < \varepsilon \). Moreover, since \( \widehat{U}_{N,i} - U_i = o_p(1) \), there exists an \( N_0 > 0 \), possibly depending on \( i \), such that \( \Pr[|\widehat{U}_{N,i} - U_i| \geq M/2] < \varepsilon/2 \) if \( N > N_0 \). Thus, under the latter condition

\[
\Pr[|\widehat{U}_{N,i}| > M] = \Pr[|U_i + (\widehat{U}_{N,i} - U_i)| > M]
\leq \Pr[|U_i| + |\widehat{U}_{N,i} - U_i| > M]
= \Pr[|U_i| + |\widehat{U}_{N,i} - U_i| > M \text{ and } |\widehat{U}_{N,i} - U_i| < M/2]
+ \Pr[|U_i| + |\widehat{U}_{N,i} - U_i| > M \text{ and } |\widehat{U}_{N,i} - U_i| \geq M/2]
\leq \Pr[|U_i| > M/2] + \Pr[|\widehat{U}_{N,i} - U_i| \geq M/2] < \varepsilon.
\]

Now for \( N > N_0 \) and arbitrary \( \xi > 0 \),

\[
\Pr[|\gamma_N(\widehat{U}_{N,i}) - \gamma_0(\widehat{U}_{N,i})| > \xi]
= \Pr[|\gamma_N(\widehat{U}_{N,i}) - \gamma_0(\widehat{U}_{N,i})| > \xi \text{ and } |\widehat{U}_{N,i}| \leq M]
+ \Pr[|\gamma_N(\widehat{U}_{N,i}) - \gamma_0(\widehat{U}_{N,i})| > \xi \text{ and } |\widehat{U}_{N,i}| > M]
\leq \Pr[\sup_{|u| \leq M} |\gamma_N(u) - \gamma_0(u)| > \xi] + \Pr[|\widehat{U}_{N,i}| > M]
\leq \Pr[\sup_{|u| \leq M} |\gamma_N(u) - \gamma_0(u)| > \xi] + \varepsilon,
\]

39
hence by (74), \( \limsup_{N \to \infty} \Pr \left[ \left| \hat{\gamma}_N(U_{N,i}) - \gamma_0(U_{N,i}) \right| > \xi \right] \leq \varepsilon \), which by the arbitrariness of \( \varepsilon \) and \( \xi \) implies that
\[
\left| \hat{\gamma}_N(U_{N,i}) - \gamma_0(U_{N,i}) \right| = o_p(1)
\]
for each \( i \). Similarly, it follows that
\[
\left| \hat{\gamma}_N(U_i) - \gamma_0(U_i) \right| = o_p(1)
\]
for each \( i \). Finally, it is not too hard to verify from the continuity of \( \gamma_0(u) \) on \( \mathbb{R} \) that
\[
\left| \bar{\gamma}_0(U_{N,i}) - \gamma_0(U_i) \right| = o_p(1)
\]
for each \( i \).\footnote{\text{As is well-known, convergence in probability is equivalent to a.s. convergence along a further subsequence of an arbitrary subsequence. See for example Bierens (2004, Ch.6). Let \( N_k \) be such a further subsequence, so that \( \hat{U}_{N,i} - U_i = o_p(1) \) is equivalent to \( \hat{U}_{N_k,i} - U_i \overset{a.s.}{\to} 0 \) as \( k \to \infty \). With \( \{\Omega, P, \mathcal{F}\} \) the probability space involved, this a.s. convergence result is equivalent to the statement that there exists a null set \( N_0 \in \mathcal{F} \), i.e., \( P(N_0) = 0 \), such that \( \lim_{k \to \infty} \hat{U}_{N_k,i}(\omega) = U_i(\omega) \) for all \( \omega \in \Omega \setminus N_0 \). Since clearly \( P(\{\omega \in \Omega : |U_i(\omega)| = \infty\}) = 0 \), we may choose \( N_0 \) such that \( |U_i(\omega)| < \infty \) for all \( \omega \in \Omega \setminus N_0 \). Then it follows from the continuity of \( \gamma_0(u) \) on \( \mathbb{R} \) that \( \lim_{k \to \infty} \gamma_0(U_{N_k,i}(\omega)) = \gamma_0(U_i(\omega)) \) for all \( \omega \in \Omega \setminus N_0 \), which implies that \( \gamma_0(U_{N,i}) = \gamma_0(U_i) + o_p(1) \).}}

Combining these three results it follows that \( \hat{\gamma}_N(U_{N,i}) - \gamma_0(U_i) = o_p(1) \) for each \( i \in \mathbb{N} \).

12.4 Computation of the SNP Laplace transform

In this section it will be shown how to compute the Laplace transform
\[
L(t|\pi_n\delta) = \int_0^\infty \exp(tw)f(w|\pi_n\delta)dw
\]
exactly for the case
\[
f(w|\pi_n\delta) = \exp(-w)\frac{(1 + \sum_{m=1}^n \delta_m \rho_m(w))^2}{1 + \sum_{k=1}^n \delta_k^2}, \quad w > 0
\]
where the \( \rho_m(w) \)'s are the Laguerre polynomials. The idea is to write each \( \rho_m(w) \) as a regular polynomial of order \( m \),
\[
\rho_m(w) = \sum_{\ell=0}^m \alpha_{\ell,m}w^\ell,
\]
where by (35),
\[
\alpha_{\ell,m} = \binom{m}{\ell} \frac{(-1)^\ell}{\ell!}
\]
and store these coefficients in a \((n+1) \times (n+1)\) lower triangular matrix \(A_{n+1}\), with upper-left element 1. Next, let denote
\[
\delta_{(n+1)} = (1, \delta_1, \delta_2, \ldots, \delta_n)
\]
and
\[
\xi(\pi_n\delta) = (\xi_0(\pi_n\delta), \xi_1(\pi_n\delta), \ldots, \xi_n(\pi_n\delta))
\]
\[
= \frac{1}{\sqrt{\delta_{(n+1)}^T A_{n+1} \delta_{(n+1)}}}
\]
Then we can write
\[
f(w|\pi_n\delta) = \exp(-w) \left( \sum_{\ell=0}^{n} \xi_\ell(\pi_n\delta) w^\ell \right)^2
\]
\[
= \sum_{m=0}^{n} \sum_{k=0}^{n} \xi_m(\pi_n\delta) \xi_k(\pi_n\delta) w^{k+m} \exp(-w)
\]
hence
\[
L(t|\pi_n\delta)
\]
\[
= \sum_{m=0}^{n} \sum_{k=0}^{n} \xi_m(\pi_n\delta) \xi_k(\pi_n\delta) \int_{0}^{\infty} w^{k+m} \exp(-(1+t)w)dw
\]
\[
= \sum_{m=0}^{n} \sum_{k=0}^{n} \xi_m(\pi_n\delta) \xi_k(\pi_n\delta) \frac{1}{(1+t)^{k+m+1}} \int_{0}^{\infty} w^{k+m} \exp(-w)dw
\]
\[
= \sum_{m=0}^{n} \sum_{k=0}^{n} \xi_m(\pi_n\delta) \xi_k(\pi_n\delta) \frac{(k+m)!}{(1+t)^{k+m+1}}
\]
\[
= \xi(\pi_n\delta)' B_{n+1}(t) \xi(\pi_n\delta)
\]
where \(B_{n+1}(t)\) is the \((n+1) \times (n+1)\) with typical element \((k+m)!(1+t)^{k+m+1}\).

Note the integral in the second equality is known as the \(\Gamma(k+m+1)\) Gamma function, satisfying \(\Gamma(1) = 1, \Gamma(k+m+1) = (k+m)\Gamma(k+m)\), hence
\[
\int_{0}^{\infty} w^{k+m} \exp(-w)dw = (k+m)!
\]
References


Chen, X. (2007), ”Large sample sieve estimation of semi-nonparametric models”. Chapter 76 in J. Heckman and E. Leamer (eds.), *Handbook of

19Downloadable from http://www.personal.psu.edu/hxb11/HANDBOOK_CHAPTER.PDF


Li, Q. (1996), ”Estimating a stochastic production frontier when the adjusted error is symmetric,” *Economics Letters* 52, 221-228.


