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

June 13, 2020

Abstract

The stochastic frontier (SF) model is in essence a linear or nonlinear regression model with an additive constant term \( \alpha_0 \) and a composite error term, \( V - W \), say, where \( V \) is an idiosyncratic error term which is usually assumed to be \( \mathcal{N}(0, \sigma^2_0) \) distributed, and \( W \) is a nonnegative random component that captures the production inefficiency in the model. In most applications the response function involved is the log of a parametric production function. Under standard regression conditions the parameters of this log-production function and the regression intercept \( \mu_0 = \alpha_0 - E[W] \) in the SF model can be estimated consistently by linear or nonlinear least squares without assuming a parametric specification of the distribution of \( W \), so the remaining questions we will answer in this paper are threefold: (1) Under what conditions are the parameters \( \alpha_0, \sigma^2_0 \) and the distribution of \( W \) semi-nonparametrically identified? (2) How can we estimate \( \alpha_0, \sigma^2_0 \) and the distribution of \( W \) semi-nonparametrically? (3) How can we semi-nonparametrically estimate the technical efficiency index \( E[\exp(-W)|V - W = \varepsilon] \) for an individual firm?

*This is a paper in progress. Please do not quote without permission of the authors.
†Professor Emeritus of Economics. Since I no longer have an office at PennState, please address correspondence by email only to hbierens@psu.edu.
‡Email: ecdhpl@ccu.edu.tw
1. Introduction

Since the pioneering works of Aigner et al. (1977, ALS hereafter) and Meeusen 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

\[
Y = \alpha_0 + X' \beta_0 + V - W, \quad (1.1)
\]

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.

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

\[
\ln(Q) = \alpha_0 + \beta_{0,1} \ln(K) + \beta_{0,2} \ln(L) + V - W, \quad (1.2)
\]

where \(\bar{Q} = \exp(V) \exp(\alpha_0 K^{\beta_0,1} L^{\beta_0,2})\) is the production frontier of an individual firm, \(\exp(-W) = Q/\bar{Q}\) 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 objectives of this paper are threefold. First, it will be shown that the SF model is semi-nonparametrically identified as long as the distribution $F(w)$ of $W$ is confined to the collection of distribution functions satisfying $F(w) > 0$ for all $w > 0$. Second, we 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. Third, in the current literature on semi-parametric SF models the consistent estimation of the technical efficiency (TE) index $E[\exp(-W)|V - W = \varepsilon]$ is still an unsolved problem, which will be solved in the current paper.

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 authors 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

\footnotesize
$^1$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.

$^2$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.

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.

The plan of the paper is as follows. In section 2 we summarize the standard conditions for the strong consistency of the OLS estimators of $\beta_0$ and the intercept $\mu_0 = \alpha_0 - E[W]$ in SF model (1.1). In section 3 we show that under a regularity condition the constant $\alpha_0$, the variance $\sigma_0^2$ of $V$ and the distribution of $W$ are semi-nonparametrically identified, regardless the type of the distribution of $W$. In section 4 we show how to estimate the pseudo-Laplace transform

$$E[\exp(-t(W - V - \alpha_0))], \ t \geq 0,$$

uniformly strongly consistent. The latter results will be verified by a numerical example, in section 5. In section 6 we show how to model the density of $W$ and its Laplace transform semi-nonparametrically, using the results in Bierens (2014a). On the basis of a further elaboration of the numerical example we show in section 7 how the SNP density and its Laplace transform fit the true density and its Laplace transform, respectively.

In section 8 we propose the SNP objective function, and in section 9 we show, after adding a penalty function to the previous objective function, that the SF model is semi-nonparametrically identified if the distribution of $W$ is absolutely continuous. In section 10 we show how to estimate the parameters $\alpha_0, \sigma_0^2 = \text{var}(V)$ and the density $f_0(w)$ in SF model (1.1) consistently via a sieve estimation approach. In section 11 we show how to estimate the technical efficiency index.
In section 12 we plan to present the sieve estimation results for the numerical example. In section 13 we plan to apply our SNP approach to estimate a wage frontier model for men and the impact of human capital on the industrial wage distributions in Taiwan. Finally, in section 13 we plan to make some concluding remarks.

This paper comes with two appendixes. Appendix A contains the proofs of some lemmas and theorems, and Appendix B is devoted to computational issues. In due course these appendixes will be detached in a separate online supplement. Moreover, the figures in sections 5, 7 and 9 are displayed in an online file only, but they will be included in the paper upon completion.

2. OLS estimation

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 are jointly i.i.d. as $(X, V, W)$ across observations. The variables $Y_i$ and $X_i$ are observed for $i = 1, 2, ..., N$.

(b) $X \in \mathbb{R}^p$, $E[X'X] < \infty$ and $\Sigma = \text{Var}(X)$ is nonsingular.\(^5\)

(c) $V$ is distributed as $\mathcal{N}(0, \sigma_0^2)$.

(d) $\Pr[W < 0] = 0$ and $E[W^2] < \infty$.

(e) $\Pr[W \leq w] > 0$ for all $w > 0$.

Now write the SF model (1.1) as

\[
Y_i = \alpha_0 - E[W] + X_i'\beta_0 + V_i - (W_i - E[W])
\]

\[
= \mu_0 + X_i'\beta_0 + U_i, \quad i = 1, 2, ..., N, \tag{2.1}
\]

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 $E[U_i^2] = \sigma_0^2 + \sigma_W^2 < \infty$, where $\sigma_W^2$ is the variance of $W$, the parameters $\mu_0$ and $\beta_0$ in model (2.1) can be estimated strongly consistent and asymptotic normally by OLS. For the purpose of references these well-known properties are listed here:

\(^4\)http://www.personal.psu.edu/hxb11/SNP_SF_PLOTS.PDF

\(^5\)The latter excludes that one of the components of $X$ is a constant, as motivated by the Cobb-Douglas example (1.2).
The OLS estimator \( \hat{\beta}_N \) of \( \beta_0 \) satisfies
\[
E[\hat{\beta}_N] = \beta_0, \quad \hat{\beta}_N \xrightarrow{a.s.} \beta_0, \quad \text{and} \quad \sqrt{N}(\hat{\beta}_N - \beta_0) \xrightarrow{d} N_p (0, (\sigma_0^2 + \sigma_W^2)\Sigma^{-1}) \tag{2.2}
\]
as \( N \to \infty \), and the OLS estimator
\[
\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{\beta}_N X_i) \tag{2.3}
\]
of \( \mu_0 \) satisfies
\[
E[\hat{\mu}_N] = \mu_0, \quad \hat{\mu}_N \xrightarrow{a.s.} \mu_0 = \alpha_0 - E[W] \tag{2.4}
\]
as \( N \to \infty \). Moreover,
\[
\hat{s}_N^2 = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{\beta}_N X_i - \hat{\mu}_N)^2 \xrightarrow{a.s.} \sigma_0^2 + \sigma_W^2 = \sigma_0^2 + E[W^2] - (E[W])^2. \tag{2.5}
\]

3. Identification: The general case

Suppose that next to the linear SF model (1.1) there exists an alternative model
\[
Y = \alpha + X'\beta_0 + V_* - W_*,
\]
where \( V_* \sim N(0, \sigma^2) \), \( W_* \) is a nonnegative random variable and \( X, V_* \) and \( W_* \) are independent, such that both models are observational equivalent, in the sense that they generate the same conditional distribution of \( Y \) given \( X \). This is only possible if \( \alpha_0 + V - W \sim \alpha + V_* - W_* \), which implies that for all \( t > 0 \),
\[
E[\exp(-t(W - V - \alpha_0))] \equiv E[\exp(-t(W^* - V^* - \alpha))].
\]

For \( t > 0 \) we can write, by the independence of \( V \) and \( W \),
\[
E[\exp(-t(W - V - \alpha_0))] = \exp(t.\alpha_0) \exp(\sigma_0^2 t^2/2)E[\exp(-t.W)]
\]
where \( \exp(\sigma_0^2 t^2/2) = E[\exp(t.V)] \) is the well-known moment generating function of \( V \sim N(0, \sigma_0^2) \), and \( E[\exp(-t.W)] \) is known as the Laplace transform of \( W \). Similarly, we have
\[
E[\exp(-t(W^* - V^* - \alpha))] = \exp(t.\alpha) \exp(\sigma^2 t^2/2)E[\exp(-t.W_*)]
\]
\[\text{Here and in the sequel "a.s." stands for "almost surely", which means that the property involved holds with probability 1. Thus, } \hat{\beta}_N \xrightarrow{a.s.} \beta_0 \text{ means } \Pr[\lim_{N \to \infty} \hat{\beta}_N = \beta_0] = 1.\]
Thus, for $t > 0$,

$$\exp(\alpha t + \sigma^2 t^2/2) \cdot E[\exp(-tW_*)] \equiv \exp(\alpha_0 t + \sigma_0^2 t^2/2) \cdot E[\exp(-tW)], \quad (3.1)$$

hence

$$E[\exp(-tW)] \equiv \exp(t(\alpha - \alpha_0) + (\sigma^2 - \sigma_0^2)t^2/2) \cdot E[\exp(-tW_*), \quad (3.2)$$

$$E[\exp(-tW_*)] \equiv \exp(t(\alpha_0 - \alpha) + (\sigma^2 - \sigma_0^2)t^2/2) \cdot E[\exp(-tW)]. \quad (3.3)$$

To prove that $\sigma^2 = \sigma_0^2$, suppose in first instance that $\sigma^2 > \sigma_0^2$. Then by (3.2),

$$E[\exp(-tW)] \geq \exp \left( t(\alpha - \alpha_0) + (\sigma^2 - \sigma_0^2)t^2/2 \right) \cdot E[\exp(-tW \cdot I(W_* \leq K))]$$

$$\geq \exp \left( t(\alpha - \alpha_0 - K) + (\sigma^2 - \sigma_0^2)t^2/2 \right) \cdot \Pr[W_* \leq K], \quad (3.6)$$

where $I(.)$ is the well-known indicator function and $K$ is chosen so large that $\Pr[W_* \leq K] > 0$. 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, (3.3) implies that $\sigma^2 < \sigma_0^2$ is not possible either, so that $\sigma^2 = \sigma_0^2$.

Thus, (3.2) and (3.3) now become

$$E[\exp(-tW)] \equiv \exp(t(\alpha - \alpha_0)) \cdot E[\exp(-tW_*), \quad (3.4)$$

$$E[\exp(-tW_*)] \equiv \exp(t(\alpha_0 - \alpha)) \cdot E[\exp(-tW)], \quad (3.5)$$

respectively.

Suppose first that $\alpha_0 > \alpha$. Then it follows from (3.5) that

$$E[\exp(-tW_*)] \geq \exp(t(\alpha - \alpha_0)) \cdot E[\exp(-tW) \cdot I(W \leq (\alpha_0 - \alpha)/2)]$$

$$\geq \exp(t(\alpha - \alpha)/2) \cdot \Pr[W \leq (\alpha_0 - \alpha)/2], \quad (3.6)$$

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.

Next, suppose that $\alpha_0 < \alpha$. Then by (3.5), similar to (3.6), we have the inequality

$$E[\exp(-tW)] \geq \exp(t(\alpha - \alpha_0)/2) \cdot \Pr[W_* \leq (\alpha - \alpha_0)/2],$$
so that by the same argument as before, $\alpha_0 < \alpha$ is not possible if $\Pr[W_* \leq (\alpha - \alpha_0)/2] > 0$. However, the latter condition is not guaranteed. Thus, at this point we can only conclude that $\gamma = \alpha - \alpha_0 \geq 0$, so that (3.5) reads

$$E[\exp(-t.W_*)] \equiv E[\exp(-t.(W + \gamma))], \forall t > 0. \quad (3.7)$$

As is well-known, two nonnegative random variables have the same distribution if and only if their Laplace transforms are equal. Thus, (3.7) implies that $W_* \sim W + \gamma$. This case corresponds to the following version of SF model (1.1):

$$Y = \alpha + X'\beta_0 + V - W_* = \alpha_0 + \gamma + X'\beta_0 + V - (W + \gamma). \quad (3.8)$$

Of course, $\gamma$ cancels out in (3.8), so that the latter model is observational equivalent to (1.1). Only if we impose (or assume) the condition

$$\Pr[W_* \leq w] > 0 \text{ for all } w > 0 \quad (3.9)$$

can we conclude that $\alpha = \alpha_0$, and then by the equality

$$E[\exp(-t.W_*]) = E[\exp(-t.W)]$$

for all $t > 0$ it follows that $W_* \sim W$.

Summarizing, the following identification result has been shown.

Theorem 1.
(a) Let $\mathcal{F}_+$ 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 (1.1). Suppose that the latter model is observational equivalent to the alternative model

$$Y = \alpha + X'\beta_0 + V - W_*,$$

where $X$, $V$ and $W_*$ are independent, $V_* \sim \mathcal{N}(0, \sigma^2)$, and the c.d.f. $F$ of $W_*$ belongs to $\mathcal{F}_+$. Then under Assumption 1, $\alpha = \alpha_0$, $\sigma^2 = \sigma_0^2$ and $F = F_0$.

(b) However, if $W_*$ is merely nonnegative, without the condition that its c.d.f. $F$ belongs to $\mathcal{F}_+$, then we can only conclude that $\sigma^2 = \sigma_0^2$, $\alpha = \alpha_0 + \gamma$ and $W_* \sim W + \gamma$ for some $\gamma \geq 0$, hence $F(w) = F_0(w - \gamma)$. C.f. (3.8).

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 1(a) applies to nonnegative discrete, continuous and mixed discrete-continuous distributions for $W$.

In the case (3.8), with the distribution of $W$ confined to the class of absolutely continuous distributions, there is a way to force $\gamma = 0$, so that then the SF model (1.1) is semi-nonparametrically identified. See section 9.
4. Estimation of the pseudo-Laplace transform of \( W - V - \alpha_0 \)

4.1. Linear SF models

Recall from the linear SF model (1.1) that

\[
X'_i(\hat{\beta}_N - \beta_0) - Y_i + ||X_i|| = X'_i(\hat{\beta}_N - \beta_0) + ||X_i|| + W_i - V_i - \alpha_0. \tag{4.1}
\]

The reason for adding the term \( ||X_i|| \) to both sides of this equation is that by (2.2),

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

for sufficient large \( N \). In particular, (2.2) 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 \( c > 0 \),

\[
\sup_{t \in [0, c]} \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t.(X'_i\hat{\beta}_N - Y_i + ||X_i||) \right) \right|
\]

\[
- \left| E[\exp (-t.(X'\xi + ||X|| + W - V - \alpha_0))]|_{\xi = \hat{\beta}_N - \beta_0} \right|
\]

\[
\leq \sup_{||\xi|| \leq 1, t \in [0, c]} \left| \frac{1}{N} \sum_{i=1}^{N} \exp \left( -t.(X'_i\xi + ||X_i|| + W_i - V_i - \alpha_0) \right) \right|
\]

\[
- \left| E[\exp (-t.(X'\xi + ||X|| + W - V - \alpha_0))]|_{\xi = \hat{\beta}_N - \beta_0} \right| \overset{a.s.}{\to} 0, \tag{4.2}
\]

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

\[
\left| E[\exp (-t.(X'\xi + ||X|| + W - V - \alpha_0))]|_{\xi = \hat{\beta}_N - \beta_0} \right|
\]

\[
- \left| E[\exp (-t.(||X|| + W - V - \alpha_0))]|_{\xi = \hat{\beta}_N - \beta_0} \right|
\]

\[
\leq E[\exp (-t.(X'\xi + ||X||)) - \exp(-t.||X||)|_{\xi = \hat{\beta}_N - \beta_0}
\]

\[
\times E[\exp (-t.(W - V - \alpha_0))]
\]

\[
\leq E[\exp(-t.X'\xi) - 1] \cdot \exp(-t.||X||)|_{\xi = \hat{\beta}_N - \beta_0}
\]

\[
\times E[\exp (-t.(W - V - \alpha_0))]
\]

\[
\leq ||\hat{\beta}_N - \beta_0||t.E[||X||] \exp(t.\alpha_0) \exp(\sigma_0^2t^2/2)E[\exp(-t.W)],
\]

9
where the last inequality follows from the mean value theorem. It follows now straightforwardly from \((2.2)\) that

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

\text{(4.3)}

Since

\[
\sup_{t \in [0,c]} \left[ \frac{1}{N} \sum_{i=1}^{N} \exp(-t.||X_i||) - E[\exp(-t.||X||)] \right] \xrightarrow{a.s.} 0
\]

\text{(4.4)}
as well, it follows from \((4.2)\), \((4.3)\) and \((4.4)\) that the following result hold.

\textbf{Lemma 1.} Denote

\[
Z_{i,N} = X_i^{\sim} \hat{\beta}_N - Y_i.
\]

\text{(4.5)}

Under Assumption 1, and for the linear SF model \((1.1)\),

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

\text{(4.6)}

for any constant \(c > 0\).

The first result \((4.8)\) in the following key theorem follows now trivially from Lemma 1.

\textbf{Theorem 2.} Denote

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

\text{(4.7)}

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

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

\text{(4.8)}
Moreover, similar to (4.8) it can be shown that for any constant \( c > 0 \),
\[
\sup_{t \in [0, c]} \left| \Phi_N(t) - a_0 - \sigma_0^2 t - d \ln \left( E[\exp(-t.\mathcal{W})] \right) \right| \overset{a.s.}{\to} 0,
\]
\[
\sup_{t \in [0, c]} \left| \Phi''_N(t) - \sigma_0^2 - d^2 \ln \left( E[\exp(-t.\mathcal{W})] \right) \right| \overset{a.s.}{\to} 0,
\]
hence
\[
\sup_{t \in [0, c]} \left| \Phi'_N(t) - t.\Phi''_N(t) - a_0 - d \ln \left( E[\exp(-t.\mathcal{W})] \right) \right| \overset{a.s.}{\to} 0,
\]
\[
+ t \left( d^2 \ln \left( E[\exp(-t.\mathcal{W})] \right) \right) \overset{a.s.}{\to} 0.
\]

**Proof.** The proofs of (4.9) and (4.10) are not too hard but rather tedious, and will be given in Appendix A. \( \blacksquare \)

Since the Laplace transform \( E[\exp(-t.\mathcal{W})] \) is unknown, Theorem 2 is not directly applicable to estimate \( a_0 \) and \( \sigma_0^2 \). However, we will show in section 7 below, for the case that \( \mathcal{W} \) has an absolutely continuous distribution, how to model any density on \( [0, \infty) \) and its corresponding Laplace transform seminonparametrically, so that the results in Theorem 2 can be converted to SNP versions. The latter can then be used to construct an SNP objective function to estimate \( a_0, \sigma_0^2 \) and the density of \( \mathcal{W} \) consistently, via a sieve estimation approach.

### 4.2. Nonlinear SF models

The linearity of the SF model (1.1) is convenient but not essential for the results in Theorem 2. If the model is nonlinear, say
\[
Y = a_0 + g(X, \beta_0) + V - W \tag{4.12}
\]
= \( \mu_0 + g(X, \beta_0) + V - (W - E[W]) \), \( \mu_0 = a_0 - E[W] \),
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 Theorem 2 carries over for \( Z_{i,N} = g(X_i, \hat{\beta}_N) - Y_i \) and with \( ||X_i|| \) replaced by \( \sup_{\beta \in \mathbb{R}^p: ||\beta - \beta_0|| \leq 1} ||g(X_i, \beta)/\partial \beta'|| \). With these modifications, all the results in this paper carry over to nonlinear SF models.
5. Numerical example: Part 1

5.1. Test model and data

In order to see how the results in Theorem 2, and later on also our SNP estimation procedure, work in practice, we have generated a random sample of size $N = 1000$ from the data-generating process

$$Y = 1 + 0.8X_1 + 0.3X_2 + V - W,$$

which corresponds to the log of a Cobb-Douglas production function with increasing returns to scale. The variables involved are generated as follows:

$$X_1 = \sqrt{5}U_1, \quad X_2 = \frac{3U_2 - U_1}{\sqrt{2}}, \quad V = 0.5U_3,$$

where $(U_1, U_2, U_3) \sim \mathcal{N}_3(0, I_3)$,

and $W$ is i.i.d. as $\text{Gamma}(3, \gamma_0)$, i.e., the density of $W$ is

$$f_0(w) = \frac{w^2 \exp(-w/\gamma_0)}{2\gamma_0^3}, \quad \gamma_0 > 0, \quad w \geq 0. \quad (5.1)$$

where the parameter $\gamma_0$ is chosen such that $\text{var}(W) = \text{var}(V) = 0.25$, which is the case for $3\gamma_0^2 = 0.25$, hence $\gamma_0 = 1/(2\sqrt{3})$. Note that $W$ can be generated as $W \sim \gamma_0 \sum_{j=1}^{3} \ln(1/(1 - U_j^*))$, where the $U_j^*$’s are random drawings from the uniform $[0, 1]$ distribution.

Thus, the true parameters are

$$\beta_0 = (\beta_{0,1}, \beta_{0,2})' = (0.8, 0.3)', \quad \alpha_0 = 1, \quad \sigma_0^2 = \text{var}(V) = \text{var}(W) = 0.25,$$

and $\text{var}(X_1) = \text{var}(X_2) = 5$, $\text{cov}(X_1, X_2) = -2.5$, so that $\text{corr}(X_1, X_2) = -0.5$. The latter allows for a realistic amount of collinearity. Moreover, $E[W] = 3\gamma_0 = 0.5\sqrt{3}$, so that the true regression intercept is

$$\mu_0 = \alpha_0 - E[W] = 1 - 0.5\sqrt{3} \approx 0.1339746.$$

\footnote{More generally, the characteristic function of the $\text{Gamma}(\alpha, \beta)$ distribution is $(1 - \beta i t)^{-\alpha}$, hence, if $\alpha \in \mathbb{N}$ then a random drawing from the $\text{Gamma}(\alpha, \beta)$ distribution is distributed as the sum of $\alpha$ independent random drawings from the distribution with characteristic function $(1 - \beta i t)^{-1}$, i.e., the exponential distribution with density $g(w) = \beta^{-1} \exp(-w/\beta)$, c.d.f. $G(w) = 1 - \exp(-w/\beta)$ and inverse $G^{-1}(u) = \beta \ln(1/(1 - u))$. Thus, with $U_j^*, \ j = 1, 2, ..., \alpha$, random drawings from the uniform $[0, 1]$ distribution, $\beta \sum_{j=1}^{\alpha} \alpha \ln(1/(1 - U_j^*)) \sim \text{Gamma}(\alpha, \beta)$.}
The OLS estimation results for \( N = 1000 \) are presented in Table 5.1.

**Table 5.1. OLS estimation results for \( N = 1000 \)**

<table>
<thead>
<tr>
<th>True parameters</th>
<th>OLS estimates</th>
<th>Standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{0,1} = 0.8 )</td>
<td>0.7851430</td>
<td>0.01097</td>
</tr>
<tr>
<td>( \beta_{0,2} = 0.3 )</td>
<td>0.2882862</td>
<td>0.01094</td>
</tr>
<tr>
<td>( \mu_0 = 0.1339746 )</td>
<td>0.1343781</td>
<td>0.02293</td>
</tr>
<tr>
<td>( \text{var}(V - W) = 0.5 )</td>
<td>0.52492807</td>
<td></td>
</tr>
</tbody>
</table>

with \( R^2 = 0.8376 \).

### 5.2. Test of Theorem 2

Our approach hinges crucially on the results of Theorem 2, so we will now verify how \( \Upsilon_N(t), \Upsilon_N'(t) \) and \( \Upsilon_N(t) - t.\Upsilon_N'(t) \) compare with their a.s. limit.

First note from the well-known moment generating function of the Gamma density (5.1) that its Laplace transform takes the form

\[
L_0(t) = \int_0^\infty \exp(-t.w) f_0(w) \, dw = (1 + \gamma_0 t)^{-3} = \left(1 + t/(2\sqrt{3})\right)^{-3}, \quad t \geq 0. \quad (5.2)
\]

and thus

\[
\ln(L_0(t)) = -3 \ln \left(1 + t/(2\sqrt{3})\right), \quad (5.3)
\]

\[
d\ln(L_0(t))/dt = -\frac{1}{2} \sqrt{3} \left(1 + t/(2\sqrt{3})\right)^{-1}, \quad (5.4)
\]

\[
d^2 \ln(L_0(t))/dt^2 = \frac{1}{4} \left(1 + t/(2\sqrt{3})\right)^{-2}. \quad (5.5)
\]

Hence, in the present case the result (4.8) in Theorem 2 reads

\[
\Upsilon_N(t) \xrightarrow{a.s.} \Upsilon(t) = t + 0.25t^2/2 - 3 \ln \left(1 + t/(2\sqrt{3})\right),
\]

uniformly on \([0, c]\) for any \( c > 0 \).

Moreover, denoting

\[
\sigma_N^2(t) = \Upsilon_N'(t) - \frac{1}{4} \left(1 + t/(2\sqrt{3})\right)^{-2}, \quad (5.6)
\]

\[
\alpha_N(t) = \Upsilon_N'(t) - t.\Upsilon_N''(t) + \frac{1}{2} \sqrt{3} \left(1 + t/(2\sqrt{3})\right)^{-1}
\]

\[
+ \frac{1}{4} t \left(1 + t/(2\sqrt{3})\right)^{-2}, \quad (5.7)
\]
the results (4.10) and (4.11) read in the present case as
\[
\sup_{t \in [0, c]} |\sigma^2_N(t) - \sigma^2_0| \xrightarrow{a.s.} 0, \text{ where } \sigma^2_0 = 0.25,
\]
\[
\sup_{t \in [0, c]} |\alpha_N(t) - \alpha_0| \xrightarrow{a.s.} 0, \text{ where } \alpha_0 = 1,
\]
respectively.

In Figures 5.1 and 5.2 we compare $\Upsilon_N(t)$ for $N = 1000$ and with its limit $\Upsilon(t)$ for $c = 1$ and $c = 5$, respectively.

- Insert
  - Figure 5.1. $\Upsilon_N(t)$ (solid line) compared with its limit $\Upsilon(t)$ (dotted line) on $[0, 1]$
  - Figure 5.2. $\Upsilon_N(t)$ (solid line) compared with its limit $\Upsilon(t)$ (dotted line) on $[0, 5]$

It appears that $\Upsilon_N(t)$ systematically underestimates $\Upsilon(t)$, with gap becoming wider with the value of $t$. But on $[0, 1]$ the fit is quite close. Therefore, $c$ should not be chosen too large.

Next, in Figures 5.3 and 5.4 we compare $\sigma^2_N(t)$ with its limit $\sigma^2_0 = 0.25$ for $c = 1$ and $c = 5$, respectively.

- Insert
  - Figure 5.3. $\sigma^2_N(t)$ (solid line) compared with $\sigma^2_0 = 0.25$ (dotted line) on $[0, 1]$
  - Figure 5.4. $\sigma^2_N(t)$ (solid line) compared with $\sigma^2_0 = 0.25$ (dotted line) on $[0, 5]$

In the case of Figure 5.3, $\sigma^2_N(t) \in (0.25, 0.2765)$ for $t \in [0, 1]$. However, as we see from Figure 5.4, for $t \in [1, 5]$, $\sigma^2_N(t)$ is way off from its asymptotic limit $\sigma^2_0 = 0.25$, as $\sigma^2_N(t) \rightarrow -0.015$ for $t \rightarrow 5$. The latter it likely due to the misfit of $\Upsilon_N(t)$ and $\Upsilon(t)$ on $[1, 5]$ in Figure 5.2. Therefore, if the distribution of $W$ were known then $\sup_{t \in [0, c]} |\sigma^2_N(t)| = 0.2765$ for $c \geq 1$ would be a good approximation of $\sigma^2_0$.\(^8\)

Finally, in Figures 5.5 and 5.6 we compare $\alpha_N(t)$ with its limit $\alpha_0 = 1$ for $c = 1$ and $c = 5$, respectively.

- Insert
  - Figure 5.5. $\alpha_N(t)$ (solid line) compared with $\alpha_0 = 1$ (dotted line) on $[0, 1]$
  - Figure 5.6. $\alpha_N(t)$ (solid line) compared with $\alpha_0 = 1$ (dotted line) on $[0, 5]$.\(^9\)

\(^8\)Of course, in this case $\text{var}(W) = \sigma^2_W$ is known as well, so that $\sigma^2_0$ can be recovered from (2.5), i.e., $\sigma^2_0 \approx 0.5249 - 0.25 = 0.2749$ in the case of the numerical example.

\(^9\)As said before, for the time being the Figures 5.1-5.6 are displayed in the link http://www.personal.psu.edu/hxb11/SNP_SF_PLOTS.PDF
As we see from figure 5.5, $\alpha_N(t) \in [0.9991, 1.0028]$ on $[0, 1]$, which is remarkably close to the true value $\alpha_0 = 1$, but $\inf_{0 \leq t \leq 1} \alpha_N(t) = 0.9991$ is slightly closer to $\alpha_0 = 1$ in absolute value. However, it appears from figure 5.6 that $\alpha_N(t) \to 1.757$ for $t \to 5$. Again, this pattern is likely due to the veering apart patterns of $\Upsilon_N(t)$ and $\Upsilon(t)$ on $[1, 5]$ in Figure 5.2. Thus, if the distribution of $W$ were known then $\inf_{0 \leq t \leq c} \alpha_N(t) = 0.9991$ for $c \geq 1$ would be a good fit for $\alpha_0$.

We have also done the same experiments for the case $N = 500$, and the results are very similar.

In conclusion, the results in Theorem 2 appear only useful in finite samples if $c = 1$, which therefore will be adopted in the sequel.

6. SNP modeling of densities and distribution functions 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 $S(f) = \{w \geq 0 : f(w) > 0\} \subset [0, \infty)$. The true c.d.f. $F_0(w)$ of $W$ is a member of $\mathcal{F}_{a+}$ but in addition its density $f_0(w)$ is continuous and positive on $(0, \infty)$.

Thus, each c.d.f. $F \in \mathcal{F}_{a+}$ 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.\(^{10}\) An example of the latter case is $F(w) = F_0(w - \gamma)$ for some positive constant $\gamma$, for which $f(w) = 0$ for $w < \gamma$. C.f. the case (3.8). In any case, $f(w)$ is Borel measurable.\(^{11}\)

6.1. 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

---

\(^{10}\)The latter condition excludes singular distributions.

\(^{11}\)Because by $f(w) = F'(w)$ a.e. on $[0, \infty)$, $f(w)$ is the pointwise limit of a sequence of continuous (hence Borel measurable) functions $(F(w + m^{-1}) - F(w))/m^{-1}$, $m \in \mathbb{N}$. See for example Bierens (2004, Theorem 2.4, p.40).
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 by definition makes \(L^2(\omega)\) a Hilbert space. Thus, a Hilbert space mimics the properties of a Euclidean space, which is therefore also a Hilbert space.

Moreover, under a mild condition 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., \(\langle \rho_m, \rho_n \rangle = I(m = k)\), with \(I(\cdot)\) 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) of orthonormal polynomials \(\rho_m\) of order \(m \geq 0\) that is complete in \(L^2(\omega)\). These polynomials obey the three-term recurrence relation (TTRR)

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

starting from \(\rho_0(x) = 1\) and \(\rho_1(x) = (c_1 - x)/\sqrt{c_2}\), with \(c_1 = \int x \omega(x) dx\) and \(c_2 = \int (x - c_1)^2 \omega(x) dx\), where the sequences \(a_k\) and \(b_k\) are specific for \(\omega(x)\). See for example Hamming (1973) for the general TTRR (6.1), and Bierens (2014a, Theorem 11) for the completeness of \(\{\rho_m\}_{m=0}^{\infty}\).

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 polynomials, for which \(\rho_0(x) = 1\), \(\rho_1(x) = x\), and \(a_k = \sqrt{k}, b_k = 0\) in TTRR (6.1).

The standard exponential density, \(\omega(w) = \exp(-w), \ w \geq 0\), generates the orthonormal Laguerre polynomials \(\{\rho_m\}_{m=0}^{\infty}\), for which \(\rho_0(w) = 1\), \(\rho_1(w) = 1 - w\), and \(a_k = k, b_k = 2k + 1\) in TTRR (6.1). However, the Laguerre polynomials also have the closed form expressions

\[
\rho_m(w) = \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^{\ell} w^{\ell},
\]

\(\ell\in\mathbb{N}\).

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 \(\rho_0(x) = 1\), \(\rho_1(x) = \sqrt{3} (1 - 2x)\), and \(a_k = 0.5k/\sqrt{4k^2 - 1}, b_k = 0.5\) in (6.1).

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.
6.2. 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

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

\[
= \int \left( \sqrt{f(x)} - \sqrt{\omega(x)} \sum_{m=0}^{n} \langle \rho_m, \sqrt{f/\omega} \rangle \cdot \rho_m(x) \right)^2 dx \quad (6.3)
\]

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

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

\[
\kappa_m = \langle \rho_m, \sqrt{f/\omega} \rangle = \int \rho_m(x) \sqrt{\omega(x)} \sqrt{f(x)} dx.
\]

Note that \( \kappa_m^2 < 1 \).\(^{12}\) However, \( \int f_n^*(x)dx = \sum_{m=0}^{n} \kappa_m^2 < 1 \) because by (6.3), \( \sum_{m=0}^{\infty} \kappa_m^2 = \int f(x)dx = 1 \), as is not hard to verify. This can be corrected by dividing \( f_n^*(x) \) by \( \sum_{m=0}^{n} \kappa_m^2 \), so that

\[
f_n(x) = \frac{\omega(x) \left( \sum_{m=0}^{n} \kappa_m \cdot \rho_m(x) \right)^2}{\sum_{k=0}^{n} \kappa_k^2}
\]

is a proper density function.

\(^{12}\)Because by Schwarz inequality,

\[
\left| \int \varphi_m(x) \sqrt{\omega(x)} \sqrt{f(x)} dx \right| < \sqrt{\int \varphi_m(x)^2 \omega(x) dx} \sqrt{\int f(x) dx} = ||\varphi_m|| = 1
\]
In most cases the first element \( \rho_0 \) of the sequence \( \{\rho_m\}_{m=0}^{\infty} \) is identical to 1. Then \( \kappa_0 = \int \sqrt{\omega(x)f(x)} \, dx \in (0,1) \) hence, denoting \( \delta_m = \kappa_m / \kappa_0 \) for \( m \in \mathbb{N} \), \( f_n(x) \) reads

\[
f_n(x) = \omega(x) \frac{(1 + \sum_{m=1}^{n} \delta_m \rho_m(x))^2}{1 + \sum_{k=1}^{\infty} \delta_k^2}.
\] (6.4)

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

\[
\lim_{n \to 0} \int |f(x) - f_n(x)| \, dx = 0.
\] (6.5)

This result implies, by Theorem 16 in Bierens (2014a), that \( \lim_{n \to \infty} f_n(x) = f(x) \) a.e.\(^{13}\) on the support \( S(\omega) \) of \( \omega(x) \), so that

\[
f(x) = \omega(x) \frac{(1 + \sum_{m=1}^{\infty} \delta_m \rho_m(x))^2}{1 + \sum_{k=1}^{\infty} \delta_k^2} \text{ a.e., } \sum_{k=1}^{\infty} \delta_k^2 < \infty.
\] (6.6)

However, it follows from Bierens (2014a, Theorem 16) that in general the sequence \( \{\delta_m\}_{m=1}^{\infty} \) may not be unique: there may exist possibly uncountably many sequences \( \{\delta_m\}_{m=1}^{\infty} \) for which (6.5) and (6.6) hold.\(^{14}\) On the other hand, if the support \( S(\omega) \) of \( \omega(x) \) is connected and is also the support of \( f(x) \), 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 the latter 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\(^{15}\) of (6.4) on the basis of Hermite polynomials, to approximate the bivariate error density of the latent variable equations in Heckman’s (1979) sample selection model in order to estimate this model semi-nonparametrically. Since then the univariate version of (6.4), with \( \omega(x) \) the standard normal density and the \( \rho_m(x) \)'s the Hermite polynomials, has become a popular approach in the econometric SNP literature. Exceptions (among others) are Bierens (2008), Bierens and Carvalho (2007) and Bierens and Song (2012), who use SNP densities based on Legendre polynomials, and Bierens (2014b), who uses SNP densities based on the cosine series.

\(^{13}\)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.

\(^{14}\)Due to the square in (6.6).

\(^{15}\)See Bierens (2014a, Theorem 17).
6.3. SNP densities and distribution functions 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 in the form of its Laplace transform. Therefore, the SNP density (6.4) with $x = w$, $\omega(w) = \exp(-w)$ and the $\rho_m(w)$'s the Laguerre polynomials seems to be the best option for SNP modeling of $f_0(w)$ and any competing density $f(w)$, due to the closed form expression (6.2) of $\rho_m(w)$.

In this case we can write the right-hand side of (6.6) as

$$f(w|\delta) = \exp(-w) \frac{1 + \sum_{m=1}^{\infty} \delta_m \rho_m(w))}{1 + \sum_{k=1}^{\infty} \delta_k^2} \text{ a.e. on } [0, \infty), \tag{6.7}$$

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

$$\Delta = \left\{ \delta = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}, \tag{6.8}$$

and the $\rho_m(w)$'s are the Laguerre polynomials.

The parameter space $\Delta$ will be endowed with the innerproduct

$$\langle \delta_1, \delta_2 \rangle = \sum_{m=1}^{\infty} \delta_{1,m} \delta_{2,m}, \ \delta_i = \{\delta_{i,m}\}_{m=1}^{\infty} \text{ for } i = 1, 2, \tag{6.9}$$

and associated pseudo-Euclidean norm

$$||\delta|| = \sqrt{\langle \delta, \delta \rangle} = \sqrt{\sum_{m=1}^{\infty} \delta_m^2} \tag{6.10}$$

and metric $||\delta_1 - \delta_2||$, so that $\Delta$ becomes a Hilbert space.\(^{16}\)

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

Definition 1. $\pi_n$ applied to $\delta = \{\delta_m\}_{m=1}^{\infty} \in \Delta$ as $\pi_n\delta$ replaces all the $\delta_m$'s with $m > n$ by zeros.

\(^{16}\)The latter follows from the fact that every Cauchy sequence in $\Delta$ converges to a point in $\Delta$, which is not hard to verify.
Then by (6.7), for \( n \in \mathbb{N} \),
\[
f(w | \pi_n \delta) = \exp(-w) \frac{(1 + \sum_{m=1}^{n} \delta_m \rho_m(w))^2}{1 + \sum_{k=1}^{n} \delta_k^2}, \quad w \geq 0,
\]
(6.11)
so that by (6.5),
\[
\lim_{n \to \infty} \int_{0}^{\infty} |f(w | \delta) - f(w | \pi_n \delta)| \, dw = 0.
\]
More precisely, we have
\[
\int_{0}^{\infty} |f(w | \pi_n \delta) - f(w | \delta)| \, dw \leq 2||\delta - \pi_n \delta||^2 + 4||\delta - \pi_n \delta|| \to 0 \quad (6.12)
\]
as \( n \to \infty \), where the limit follows from the fact that \( \sum_{m=1}^{\infty} \delta_m^2 < \infty \) implies \( ||\delta - \pi_n \delta||^2 = \sum_{m=n+1}^{\infty} \delta_m^2 \to 0 \), 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.** See Appendix A. \( \blacksquare \)

The SNP density (6.11) can easily be computed via the TTRR relation for Laguerre polynomials, i.e.,
\[
\rho_{k+1}(w) = \left( \frac{2k + 1 - w}{k + 1} \right) \rho_k(w) - \left( \frac{k}{k + 1} \right) \rho_{k-1}(w), \quad k \in \mathbb{N},
\]
starting from \( \rho_0(w) = 1, \rho_1(w) = 1 - w \), which is the fastest and most stable way to compute (6.11).

However, substituting the closed form expression (6.2) for \( \rho_k(w) \) and \( \rho_m(w) \) in (6.11) we can also write \( f(w | \pi_n \delta) \) as a quadratic form, i.e.,
\[
f(w | \pi_n \delta) = \frac{\xi_{n+1}(\delta)' D_{n+1}(w) \xi_{n+1}(\delta)}{\xi_{n+1}(\delta)' \xi_{n+1}(\delta)}, \quad \text{where} \quad \xi_{n+1}(\delta) = (1, \delta_1, ..., \delta_n)',
\]
(6.13)
and \( D_{n+1}(w) \) is the \( (n+1) \times (n+1) \) matrix-valued function with \( (k, m) \) element
\[
d_{k,m}(w) = \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(-1)^{i+j}}{i!j!} w^{i+j} \exp(-w)
\]
(6.14)
for \( k, m = 0, 1, ..., n \). The quadratic form (6.13) is convenient in computing the corresponding distribution function, Laplace transform and moments of \( f(w | \pi_n \delta) \).
In particular, the c.d.f. $F(w|\pi_n \delta)$ of $f(w|\pi_n \delta)$ takes the form

$$F(w|\pi_n \delta) = \frac{\xi_{n+1}(\delta)^t C_{n+1}(w) \xi_{n+1}(\delta)}{\xi_{n+1}(\delta)^t \xi_{n+1}(\delta)}$$

(6.15)

where $C_{n+1}(w) = \int_0^w D_{n+1}(x)dx$ is the $(n+1) \times (n+1)$ matrix-valued function with $(k, m)$ element

$$c_{k,m}(w) = \int_0^w d_{k,m}(x)dx$$

$$= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \binom{i+j}{i}(-1)^{i+j} \times \left(1 - \exp(-w) \left(1 + \sum_{\ell=1}^{i+j} \frac{w^\ell}{\ell!}\right)\right).$$

(6.16)

See Appendix A: Miscellaneous derivations.

Finally, note that by Lemma 2,

$$\sup_{w>0} |F(w|\pi_n \delta) - F(w|\delta)| \leq 2||\delta - \pi_n \delta||^2 + 4||\delta - \pi_n \delta|| \to 0$$

(6.17)

as $n \to \infty$.

### 6.4. Computation and convergence of moments of SNP densities

It follows straightforwardly from (6.13) and (6.14) that for $\ell \in \mathbb{N},$

$$\int_0^\infty w^\ell f(w|\pi_n \delta)dw = \frac{\xi_{n+1}(\delta)^t E_{\ell,n+1} \xi_{n+1}(\delta)}{\xi_{n+1}(\delta)^t \xi_{n+1}(\delta)},$$

(6.18)

where $E_{\ell,n+1}$ is the $(n+1) \times (n+1)$ matrix with $(k, m)$ element

$$e_{\ell,k,m} = \int_0^\infty w^\ell d_{k,m}(w)dw$$

$$= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \binom{i+j}{i}(-1)^{i+j} \int_0^\infty w^{i+j+\ell} \exp(-w)dw$$

$$= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \binom{i+j}{i} \frac{(i+j+\ell)!}{(i+j)!}(-1)^{i+j}.\quad (6.19)$$
Note that for each \( n \in \mathbb{N} \) the matrix \( E_{\ell,n+1} \) is finite, and so is (6.18):

For all \( \delta \in \Delta \) and each \( n \in \mathbb{N} \),

\[
\int_0^\infty w^\ell f(w|\pi_n \delta)dw < \infty,
\]
even if \( \int_0^\infty w^\ell f(w|\delta)dw = \infty \).

The question now arises: Does \( \int_0^\infty w^\ell f(w|\pi_n \delta)dw \) converge for \( n \to \infty \), and if so, what is its limit? The answer is given in the following lemma.

**Lemma 3.** For each \( \delta \in \Delta \) with corresponding SNP density \( f(w|\delta) \) on \([0, \infty)\) or \((0, \infty)\), and for \( \ell \in \mathbb{N} \), we have

\[
\lim_{n \to \infty} \int_0^\infty w^\ell f(w|\pi_n \delta)dw = \int_0^\infty w^\ell f(w|\delta)dw,
\]
regardless whether the latter moment is finite or not.

**Proof.** See Appendix A. \( \blacksquare \)

### 6.5. Laplace transforms

The result (6.6) 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 \([0, \infty)\).\(^{17}\) 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 \([0, \infty)\).\(^{18}\) 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.e." no longer applies after integration, it follows that in general the Laplace identity

\[
\int_0^\infty \exp(-t.w)f(w)dw \equiv \int_0^\infty \exp(-t.w)f(w|\delta)dw, \quad t > 0,
\]
holds for at least one \( \delta \in \Delta \), but possibly for uncountable many \( \delta \)'s in \( \Delta \) as well, whereas the Laplace identity

\[
\int_0^\infty \exp(-t.w)f_0(w)dw \equiv \int_0^\infty \exp(-t.w)f(w|\delta^0)dw, \quad t > 0,
\]


holds for a unique $\delta^0 \in \Delta$.

Finally, it follows trivially from Lemma 2 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 $n \to \infty$.

6.6. Computation of the SNP Laplace transform

As we will see below, in sieve estimation the infinite dimensional parameters $\delta$ involved are truncated as $\pi_n \delta$, where $n_N$ is a subsequence of $N$. Therefore, for the estimation purpose it suffices to focus on the computation of the SNP Laplace transform $L(t|\pi_n \delta)$ for some $n \in \mathbb{N}$. In particular, it follows from (6.13) and (6.14) that

$$L(t|\pi_n \delta) = \int_0^\infty \exp(-t.w)f(w|\pi_n \delta)dw = \frac{\xi_{n+1}(\delta)'B_{n+1}(t)\xi_{n+1}(\delta)}{\xi_{n+1}(\delta)'\xi_{n+1}(\delta)}$$

(6.20)

where $B_{n+1}(t)$ is the $(n+1) \times (n+1)$ matrix-valued function with $(k,m)$ element

$$b_{k,m}(t) = \int_0^\infty \exp(-t.w)a_{k,m}(w)dw$$

$$= \sum_{i=0}^k \sum_{j=0}^m \binom{k}{i} \binom{m}{j} \left(\frac{i+j}{i}\right)^{i+j+1} (\frac{1}{1+t})^i (\frac{1}{1+t})^j$$

(6.21)

See Appendix A: Miscellaneous derivations.

However, the binomial numbers in $b_{k,m}(t)$ may become too big to be computed numerically, because binomial numbers are integer valued and in most computer languages the integer size is restricted.\(^\text{19}\) Indeed, the computation of $L(t|\pi_n \delta)$ with the binomial numbers in $b_{k,m}(t)$ computed exactly appears only feasible for $n \leq 11$. On the other hand, by evaluating these three binomial numbers in log form and then taking the exp() function of their sum works for at least $n = 20$ with little loss of accuracy, as will be demonstrated in the next section, provided that the log() and exp() operations are done in double-precision.

\(^{19}\)For example, in Visual Basic 5 the maximum positive (long) integer allowed is about 2.1 billion.
7. Numerical example: Part 2

7.1. True SNP parameters $\delta^0$

As we have seen before, the Gamma density (5.1) has the series representation

$$f(w|\delta^0) = \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{0,m}\rho_m(w)}{1 + \sum_{k=1}^{\infty} \delta_{0,k}^2} \right)^2 = f_0(w) \text{ a.e.},$$

where the $\rho_m(w)$'s are the Laguerre polynomials and $\delta^0 = \{\delta_{0,m}\}_{m=1}^{\infty} \in \Delta$ is unique. In particular, it follows from Bierens (2014a, Theorem 21) that

$$\delta_{0,m} = \frac{\int_0^{\infty} \rho_m(w) \exp(-w/2) \sqrt{f_0(w)} dw}{\int_0^{\infty} \exp(-w/2) \sqrt{f_0(w)} dw}, \ m \in \mathbb{N}. \quad (7.1)$$

Using the closed form expression (6.2) for $\rho_m(w)$, and denoting $\lambda_0 = (1+1/\gamma_0)/2 = 0.5 + \sqrt{3}$, the integrals in (7.1) can be computed exactly, with results

$$\delta_{0,m} = \sum_{\ell=0}^{m} (-1)^\ell \binom{m}{\ell} \frac{\ell + 1}{\lambda_0^{\ell}}, \ m \in \mathbb{N}. \quad (7.2)$$

See Appendix A: Miscellaneous derivations.

The numerical values of the $\delta_{0,m}$'s for $m = 1, 2, ..., 20$ are tabulated in Table 7.1.

**Table 7.1.** $\delta_{0,m}$ for $m = 1, 2, ..., 20$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\delta_{0,m}$</th>
<th>$m$</th>
<th>$\delta_{0,m}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10396305</td>
<td>11</td>
<td>-0.01149065</td>
</tr>
<tr>
<td>2</td>
<td>-0.18991224</td>
<td>12</td>
<td>-0.00699196</td>
</tr>
<tr>
<td>3</td>
<td>-0.24133193</td>
<td>13</td>
<td>-0.00421785</td>
</tr>
<tr>
<td>4</td>
<td>-0.20855839</td>
<td>14</td>
<td>-0.00252602</td>
</tr>
<tr>
<td>5</td>
<td>-0.15671087</td>
<td>15</td>
<td>-0.00150352</td>
</tr>
<tr>
<td>6</td>
<td>-0.10945869</td>
<td>16</td>
<td>-0.00089019</td>
</tr>
<tr>
<td>7</td>
<td>-0.07309112</td>
<td>17</td>
<td>-0.00052464</td>
</tr>
<tr>
<td>8</td>
<td>-0.04733962</td>
<td>18</td>
<td>-0.00030796</td>
</tr>
<tr>
<td>9</td>
<td>-0.02999153</td>
<td>19</td>
<td>-0.00018013</td>
</tr>
<tr>
<td>10</td>
<td>-0.01868593</td>
<td>20</td>
<td>-0.00010502</td>
</tr>
</tbody>
</table>

with

$$\sum_{m=1}^{20} \delta_{0,m}^2 = 0.194192 \quad (7.3)$$
7.2. Fit of the SNP density of $W$

In order to check how well the corresponding SNP densities $f(w|\pi_n,\delta^0)$ fit the true Gamma density (5.1) of $W$, the latter has been compared with $f(w|\pi_n,\delta^0)$ for $w \in [0, 2.5]$ and $n = 5, 10, 15, 20$. However, only the cases $n = 5$ and $n = 10$ will be displaced, in Figures 7.1 and 7.2, respectively.

- Insert
  
  Figure 7.1. Comparison of $f_0(w) = f(w|\delta^0)$ (solid line) with $f(w|\pi_n,\delta^0)$ (dotted line) on $[0, 2.5]$ for $n = 5$
  
  Figure 7.2. Comparison of $f_0(w) = f(w|\delta^0)$ (solid line) with $f(w|\pi_n,\delta^0)$ (dotted line) on $[0, 2.5]$ for $n = 10$

It appears that for $n = 5$ the fit is poor, but for $n = 10$ the fit is excellent. Moreover, for $n = 15$ and $n = 20$ the fit is almost perfect, in the sense that the two curves are indistinguishable.

7.3. Fit of the SNP log-Laplace transform

Recall that the general SNP Laplace transform $\mathcal{L}(t|\pi_n,\delta)$ has closed-form expression (6.20), and that in the present case $\lim_{n \to \infty} \mathcal{L}(t|\pi_n,\delta^0) = \mathcal{L}(t|\delta^0) \equiv \mathcal{L}_0(t)$, where the latter is defined in (5.2). Since our approach is based on $\Psi_N(t|\delta) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\delta))$ we have verified how well $\ln(\mathcal{L}(t|\pi_n,\delta^0))$ fits $\ln(\mathcal{L}_0(t))$, for $n = 5, 10$ and $t \in [0, 5]$,

- Insert
  
  Figure 7.3. Comparison of $\ln(\mathcal{L}_0(t))$ (solid line) with $\ln(\mathcal{L}(t|\pi_n,\delta^0))$ (dotted line) on $[0, 5]$ for $n = 5$
  
  Figure 7.4. Comparison of $\ln(\mathcal{L}_0(t))$ (solid line) with $\ln(\mathcal{L}(t|\pi_n,\delta^0))$ (dotted line) on $[0, 5]$ for $n = 10^{20}$

For $n = 5$ the fit is not too bad but not good enough either, in the sense that the deviation of $\ln(\mathcal{L}(t|\pi_n,\delta^0))$ from $\ln(\mathcal{L}_0(t))$ is clearly visible. On the other hand, for $n = 10$ the fit is almost perfect. We also have looked at the cases $n = 15$ and $n = 20$, and in these cases the curves are indistinguishable. Of course, these results correspond to the fit of the SNP densities $f(w|\pi_n,\delta^0)$ to the true Gamma density (5.1).

\footnote{Again, for the time being the figures 7.1-7.4 are displayed in the link http://www.personal.psu.edu/hxb11/SNP_SF_PLOTS.PDF}
8. The SNP objective function

8.1. Theorem 2 revised

Denoting
\[ L(t|\delta) = \int_0^\infty \exp(-t.w) f(w|\delta) dw, \] (8.1)
\[ \Psi_N(t|\delta) = \Upsilon_N(t) - \ln(L(t|\delta)), \] (8.2)
\[ \Lambda(t|\delta) = \ln(L(t|\delta^0)) - \ln(L(t|\delta)), \] (8.3)
where \( \Upsilon_N(t) \) is defined in Theorem 2 and \( \delta^0 \in \Delta \) is the unique parameter such that \( L(t|\delta^0) = E[\exp(-t.W)] \), it follows trivially from Theorem 2, with \( c = 1 \), that under Assumptions 1 and 2, and for any \( \delta \in \Delta \),
\[ \sup_{t \in [0,1]} |\Psi_N(t|\delta) - \alpha_0 t - \sigma_0^2 t^2 / 2 - \Lambda(t|\delta)| \xrightarrow{a.s.} 0, \] (8.4)
\[ \sup_{t \in [\tau,1]} |\Psi''_N(t|\delta) - \alpha_0 - \Lambda''(t|\delta)| \xrightarrow{a.s.} 0, \] (8.5)
\[ \sup_{t \in [\tau,1]} |\Psi'_N(t|\delta) - t.\Psi''_N(t|\delta) - \alpha_0 - \Lambda'(t|\delta) + t.\Lambda''(t|\delta)| \xrightarrow{a.s.} 0, \] (8.6)
where \( \tau \in (0,1) \) is a small a priori chosen number and the derivatives are to \( t \).

The reason for this small \( \tau \) in (8.5) and (8.6) is that
\[ \Lambda''(0|\delta) = E[W^2] - (E[W])^2 - \int_0^\infty w^2 f(w|\delta) dw + \left( \int_0^\infty w f(w|\delta) dw \right)^2 \]
\[ = -\infty \text{ if } \int_0^\infty w^2 f(w|\delta) dw = \infty, \]
and
\[ \Lambda'(0|\delta) = \int_0^\infty w f(w|\delta) dw - E[W] = \infty \text{ if } \int_0^\infty w f(w|\delta) dw = \infty, \]
so that the results (8.5) and (8.6) will not hold for \( \tau = 0 \) in these cases.

On the other hand, since \( \Lambda(t|\delta^0) \equiv 0 \), the results in Theorem 2 imply
\[ \sup_{t \in [0,1]} |\Psi_N(t|\delta^0) - \alpha_0 t - \sigma_0^2 t^2 / 2| \xrightarrow{a.s.} 0, \] (8.7)
\[ \sup_{t \in [0,1]} |\Psi''_N(t|\delta^0) - \sigma_0^2| \xrightarrow{a.s.} 0, \] (8.8)
\[ \sup_{t \in [0,1]} |\Psi'_N(t|\delta^0) - t.\Psi''_N(t|\delta^0) - \alpha_0| \xrightarrow{a.s.} 0. \] (8.9)
8.2. Empirical objective function

The result (8.4) suggests to use the empirical objective function

$$Q_N(\alpha, \sigma^2, \delta) = \int_0^1 (\Psi_N(t|\delta) - \alpha t - \sigma^2 t^2/2)^2 \, dt$$  \hspace{1cm} (8.10)$$

in order to estimate $\alpha_0$, $\sigma^0_0$ and $\delta_0$, via a sieve estimation approach. Note that by (8.4), pointwise in $\alpha, \sigma^2, \delta$,

$$Q_N(\alpha, \sigma^2, \delta) \overset{\text{a.s.}}{\to} Q(\alpha, \sigma^2, \delta), \text{ where}$$

$$Q(\alpha, \sigma^2, \delta) = \int_0^1 (\Lambda(t|\delta) - (\alpha - \alpha_0)t - (\sigma^2 - \sigma^0_0)t^2/2)^2 \, dt.$$  \hspace{1cm} (8.11)$$

Moreover, it follows from Lemma 2 that for any pair $\delta_1, \delta_2$ in $\Delta$,

$$\sup_{t \in [0,1]} |\ln(L(t|\delta_1)) - \ln(L(t|\delta_2))| \leq \frac{2||\delta_1 - \delta_2||^2 + 4||\delta_1 - \delta_2||}{\min(L(1|\delta_1), L(1|\delta_2))}.$$  \hspace{1cm} (8.12)$$

Consequently, $Q_N(\alpha, \sigma^2, \delta)$ and its pointwise limit $Q(\alpha, \sigma^2, \delta)$ are continuous in all of their arguments.

8.3. Concentrating $\alpha$ and $\sigma^2$ out

Note that in the case of the numerical example, $\Psi_N'(t|\delta^0) - t.\Psi_N''(t|\delta^0)$ is just the function $\alpha_N(t)$ in (5.7), and as we have seen from Figures 5.5 and 5.6,

$$\overline{\alpha}_N(\delta^0) = \inf_{t \in [0,c]} (\Psi_N(t|\delta^0) - t.\Psi_N''(t|\delta^0))$$

gives an excellent approximation of $\alpha_0$, with common value for $c = 1$ and $c = 5$.

Similarly, in this case the function $\sigma^2_N(t)$ in (5.6) is just $\Psi_N''(t|\delta^0)$, and in view of Figures 5.3 and 5.4

$$\overline{\sigma}^2_N(\delta^0) = \sup_{t \in [0,c]} \Psi_N''(t|\delta^0)$$

gives a good approximation of $\sigma^0_0$, with common value for $c = 1$ and $c = 5$.

These results suggest to concentrate $\alpha$ and $\sigma^2$ out in the objective function (8.10) by replacing them by

$$\overline{\alpha}_N(\delta) = \min_{i=1,2,...,K} (\Psi_N'(i/K|\delta) - (i/K).\Psi_N''(i/K|\delta))$$  \hspace{1cm} (8.13)$$

$$\overline{\sigma}^2_N(\delta) = \max_{i=1,2,...,K} \Psi_N''(i/K|\delta)$$  \hspace{1cm} (8.14)$$
respectively, for an a priori chosen grid size \( K \), for example, let \( K = 25 \). Note that by (8.5) and (8.6),

\[
\alpha_N(\delta) \xrightarrow{a.s.} \alpha(\delta) \text{ pointwise in } \delta \in \Delta, \text{ where } \\
\alpha(\delta) = \alpha_0 + \min_{i=1,2,\ldots,K} (\Delta'(i/K|\delta) - (i/K).\Delta''(i/K|\delta)) , \\
\alpha(\delta^0) = \alpha_0
\]

and

\[
\sigma^2_N(\delta) \xrightarrow{a.s.} \sigma^2(\delta) \text{ pointwise in } \delta \in \Delta, \text{ where } \\
\sigma^2(\delta) = \sigma_0^2 + \max_{i=1,2,\ldots,K} \Delta''(i/K|\delta), \\
\sigma^2(\delta^0) = \sigma_0^2.
\]

Moreover, it follows straightforwardly from Lemma 2 that \( \alpha_N(\delta), \alpha(\delta), \sigma^2_N(\delta) \) and \( \sigma^2(\delta) \) are all (a.s.) continuous in \( \delta \).

However, the pointwise convergence results in (8.15) and (8.16) also hold uniformly in \( \delta \), i.e.,

\[
\sup_{\delta \in \Delta} |\alpha_N(\delta) - \alpha(\delta)| \xrightarrow{a.s.} 0 \text{ and } \sup_{\delta \in \Delta} |\sigma^2_N(\delta) - \sigma^2(\delta)| \xrightarrow{a.s.} 0. \tag{8.17}
\]

To see this, observe from (8.2) and Theorem 2 that

\[
\Psi''_N(t|\delta) = \Upsilon''_N(t) - \frac{d^2 \ln(\mathcal{L}(t|\delta))}{(dt)^2} \xrightarrow{N \to \infty} \Upsilon''_N(t) + \sigma_0^2 + \Delta''(t|\delta),
\]

hence

\[
\sigma^2_N(\delta) = \sigma_0^2 + \max_{i=1,2,\ldots,K} \left( \Upsilon''_N(i/K) - \lim_{N \to \infty} \Upsilon''_N(i/K) + \Delta''(i/K|\delta) \right) \\
\leq \sigma^2(\delta) + \sup_{0 \leq t \leq 1} \left| \Upsilon''_N(t) - \lim_{N \to \infty} \Upsilon''_N(t) \right| \\
\geq \sigma^2(\delta) - \sup_{0 \leq t \leq 1} \left| \Upsilon''_N(t) - \lim_{N \to \infty} \Upsilon''_N(t) \right|
\]

Of course, the reason for taking the minimum and maximum over a grid instead of the infimum and supremum over \([\tau,1]\), respectively, is merely computational convenience.

In particular, \( |\sigma^2(\delta) - \sigma^2(\delta^*)| \leq |\sum_{i=1}^{K} (\Delta'(i/K|\delta) - \Delta'(i/K|\delta^*))| \to 0 \) if \( ||\delta - \delta^*|| \to 0, \) and similar for \( \alpha(\delta), \alpha_N(\delta) \) and \( \sigma^2_N(\delta) \).

28
and thus \( \sup_{\delta \in \Delta} |\overline{\sigma}_N^2(\delta) - \overline{\sigma}^2(\delta)| \xrightarrow{a.s.} 0 \). By a similar argument it follows that \( \sup_{\delta \in \Delta} |\overline{\sigma}_N(\delta) - \overline{\sigma}(\delta)| \xrightarrow{a.s.} 0 \).

Replacing \( \alpha \) and \( \sigma^2 \) in (8.10) by \( \overline{\alpha}_N(\delta) \) and \( \overline{\sigma}_N^2(\delta) \) yield the concentrated objective function and its a.s. limit, i.e.,

\[
Q_N(\delta) = \int_0^1 (\Psi_N(t|\delta) - \overline{\sigma}_N(\delta).t - \overline{\sigma}_N^2(\delta).t^2/2)^2 \, dt \xrightarrow{a.s.} (8.18)
\]

and

\[
Q(\delta) = \int_0^1 (\Lambda(t|\delta) - (\overline{\alpha}(\delta) - \sigma_0).t - (\overline{\sigma}^2(\delta) - \sigma_0^2).t^2/2)^2 \, dt \xrightarrow{a.s.} (8.19)
\]

with \( Q(\delta^0) = 0 \).

9. SNP identification: The absolute continuous case

9.1. Identification by penalization

Now suppose that \( \delta^*_\in \Delta \) corresponds to the Laplace transform of \( W_\gamma = W + \gamma \) for some \( \gamma > 0 \). C.f. model (3.8). As follows from Theorem 1(b), this the only case for which \( \alpha_0 \) is not identified.

Note that the c.d.f. of \( W_\gamma \) is \( F(w|\delta^*_\) = \Pr[W \leq w - \gamma] = I(w > \gamma).F_0(w - \gamma), \) with density \( f(w|\delta^*_\) = I(w > \gamma).f_0(w - \gamma) \) a.e., hence,

\[
\Lambda(t|\delta^*_\) = \ln(\Lambda(t|\delta^0)) - \ln(\Lambda(t|\delta^*_\)) = \gamma.t.
\]

Consequently, (8.4) and (8.6) become

\[
\sup_{t \in [0,1]} |\Psi_N(t|\delta^*_\) - (\alpha_0 + \gamma)t - \sigma_0^2t^2/2| \xrightarrow{a.s.} 0, \quad (9.1)
\]

\[
\sup_{t \in [0,1]} |\Psi'_N(t|\delta^*_\) - t.\Lambda'(t|\delta^*_\) - (\alpha_0 + \gamma)| \xrightarrow{a.s.} 0, \quad (9.2)
\]

respectively, and thus

\[
\overline{\sigma}_N(\delta^*_\) \xrightarrow{a.s.} \overline{\sigma}(\delta^*_\) = \alpha_0 + \gamma. \quad (9.3)
\]

However, because \( \Lambda''(t|\delta^*_\) \equiv 0 \) the result (8.5) is not affected, i.e.,

\[
\overline{\sigma}_N^2(\delta^*_\) \xrightarrow{a.s.} \overline{\sigma}^2(\delta^*_\) = \sigma_0^2, \quad (9.4)
\]

as confirmed by Theorem 1(b). Hence,

\[
\overline{Q}_N(\delta^*_\) \xrightarrow{a.s.} \overline{Q}(\delta^*_\) = \overline{Q}(\delta^0) = 0.
\]
So the problem is how to restrict \( \delta \) such that, asymptotically, \( \gamma > 0 \) is not possible.

Let \( F(w|\delta) \) be a c.d.f. such that \( \inf_{w>0:F(w|\delta)>0} w = \gamma > 0 \). Then the inverse of \( F(w|\delta) \), which is defined as

\[
F^{-1}(u|\delta) = \inf_{w>0:F(w|\delta)>u} w, \quad u \in [0, 1),
\]

satisfies \( F^{-1}(0|\delta) = \gamma \). Note that by the continuity of \( F(w|\delta) \) in \( w \), (9.5) is equivalent to

\[
F^{-1}(u|\delta) = \sup_{w\geq0:F(w|\delta)\leq u} w \text{ for } u \in [0, 1).
\]

This suggests to penalize the concentrated objection function \( \overline{Q}_N(\delta) \) by

\[
\Pi(\delta) = C.F^{-1}(0|\delta)
\]

for some large positive constant \( C \), and use the penalized objective function

\[
\hat{Q}_N(\delta) = \overline{Q}_N(\delta) + \Pi(\delta)
\]

with a.s. pointwise limit

\[
\hat{Q}_N(\delta) \overset{\text{a.s.}}{\to} Q(\delta), \text{ where } Q(\delta) = \overline{Q}(\delta) + \Pi(\delta).
\]

Then for \( \delta_* \) as before, \( Q(\delta_*) = C.F^{-1}(0|\delta_*) = C.\gamma \).

Next, consider the set

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

We will now show that \( \Delta_0 = \{ \delta^0 \} \).

Clearly, the \( \delta_*'s \) for which \( f(w|\delta_*) = I(w > \gamma)f_0(w-\gamma) \) a.e. for some \( \gamma > 0 \) are excluded from \( \Delta_0 \), and so are all \( \delta \)'s for which \( F^{-1}(0|\delta) > 0 \). Thus,

\[
\delta \in \Delta_0 \text{ implies } F(w|\delta) > 0 \text{ for all } w > 0.
\]

Moreover, \( \delta \in \Delta_0 \) implies

\[
0 = \overline{Q}(\delta) = Q(\mu(\delta), \sigma^2(\delta), \delta)
= \int_0^1 (\Lambda(t|\delta) - (\mu(\delta) - \alpha_0)t - (\sigma^2(\delta) - \sigma_0^2)t^2/2)^2 \, dt.
\]

(9.10)
C.f. (8.19). The equality (9.10) implies that
\[ \Lambda(t|\delta) \equiv (\bar{\alpha}(\delta) - \alpha_0)t + (\bar{\sigma}^2(\delta) - \sigma_0^2)t^2/2 \]
on (0, 1), which by (8.3) implies that
\[ L(t|\delta^0) \equiv L(t|\delta) \exp \left( (\bar{\alpha}(\delta) - \alpha_0)t + (\bar{\sigma}^2(\delta) - \sigma_0^2)t^2/2 \right) \text{ on } (0, 1). \quad (9.11) \]

Since the left hand side of (9.11) is a Laplace transform, so is the right hand side. Therefore, (9.11) holds for all \( t \geq 0 \), because it is (our should be) well-known that if two Laplace transforms are identical on \((0, \kappa)\) for an arbitrary \( \kappa > 0 \) then the corresponding distributions are equal, hence these Laplace transforms are equal for all \( t \geq 0 \). Thus, (9.11) now reads
\[ E[\exp(-t.W)] \equiv E[\exp(-t.W_\ast)] \exp \left( (\bar{\alpha}(\delta) - \alpha_0)t + (\bar{\sigma}^2(\delta) - \sigma_0^2)t^2/2 \right) \quad (9.12) \]
on [0, \infty), where \( W \sim F(w|\delta^0) \) and \( W_\ast \sim F(w|\delta) \). Then by (9.9) and (9.12) it follows from the proof of Theorem 1(a) that \( \bar{\alpha}(\delta) = \alpha_0 \) and \( \bar{\sigma}^2(\delta) = \sigma_0^2 \), hence \( E[\exp(-t.W)] \equiv E[\exp(-t.W_\ast)] \) on \([0, \infty)\) and thus \( F(w|\delta) = F(w|\delta^0) \) for all \( w > 0 \). The latter implies, by the uniqueness of \( \delta^0 \), that \( \delta = \delta^0 \), hence \( \Delta_0 = \{\delta^0\} \).

Thus, 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 the penalized objective function satisfies \( \bar{Q}_N(\delta) \overset{a.s.}{\to} 0 \). Consequently, \( f_0(w) \) itself is semi-nonparametrically identified. Moreover, \( \alpha_0 \) and \( \sigma_0^2 \) are identified by
\[ \bar{\alpha}_N(\delta^0) \overset{a.s.}{\to} \alpha_0 \text{ and } \bar{\sigma}_N^2(\delta^0) \overset{a.s.}{\to} \sigma_0^2. \]

**9.2. How effective is the penalty function in practice?**

As we will see in the next section, the penalized objective function (9.8) will be used in the form
\[ \bar{Q}_N(\pi_n, \delta) = \bar{Q}_N(\pi_n \delta) + \Pi(\pi_n, \delta), \text{ with } \Pi(\pi_n \delta) = C.F^{-1}(0|\pi_n \delta), \]

23If not well-known, here is a quick proof. Let \( W \) and \( W_\ast \) be nonnegative random variables such that \( E[\exp(-tW)] = E[\exp(-tW_\ast)] \) for all \( t \in (0, \kappa) \). Then for such a \( t \) we can infinitely many times take the derivatives of these Laplace transform, resulting in the equalities \( E[W^m \exp(-t.W)] = E[W_\ast^m \exp(-t.W_\ast)] \) 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 \( \exp(.) \) function that \( E[\exp(i,\xi W) \exp(-t.W)] = \sum_{m=0}^{\infty} (1/m!) (i,\xi)^m E[W^m \exp(-t.W)] = E[\exp(i,\xi W_\ast) \exp(-t.W_\ast)] \). Letting \( t \downarrow 0 \) yields \( E[\exp(i,\xi W)] = E[\exp(i,\xi W_\ast)] \), which implies \( W \sim W_\ast \) by the uniqueness of characteristic functions.
where \( n \) increases to infinity with \( N \). See (10.2) below.

Moreover,
\[
\lim_{n \to \infty} F^{-1}(0|\pi_n \delta) = F^{-1}(0|\delta), \tag{9.13}
\]
as follows from the following adaptation of Lemma 3 in Bierens and Song (2012).

**Lemma 4.** For each pair \( \delta, \delta* \in \Delta \) with \( \delta \) hold fixed, and for each \( u \in [0, 1) \),
\[
F^{-1}(u|\delta*) \to F^{-1}(u|\delta) \quad \text{as} \quad ||\delta - \delta*|| \to 0.
\]

**Proof.** See Appendix A. \( \blacksquare \)

Recall that the purpose of the penalty function is to exclude the \( \delta(\gamma) \)'s for which
\[
f(w|\delta(\gamma)) = I(w > \gamma)f_0(w - \gamma) \quad \text{a.e. on} \quad [0, \infty) \tag{9.14}
\]
for some \( \gamma > 0 \), where \( f_0(w) \) is the density of the actual \( W \) in SF model (1.1).
In this case \( F^{-1}(0|\delta(\gamma)) = \gamma \), but our concern is that possibly \( F^{-1}(0|\pi_n \delta(\gamma)) = 0 \) for modest values of \( n \), despite the fact that by (9.13), \( \lim_{n \to \infty} F^{-1}(0|\pi_n \delta(\gamma)) = F^{-1}(0|\delta(\gamma)) = \gamma \).

For the Gamma density (5.1) in the numerical example the elements \( \delta_m(\gamma) \) of \( \delta(\gamma) \) can be computed exactly as
\[
\delta_m(\gamma) = \sum_{\ell=0}^{m} (-1)\ell \binom{m}{\ell} \sum_{k=0}^{\ell} \frac{k + 1}{(\ell - k)!} \lambda_0^{-k} \gamma^\ell-k, \quad m \in \mathbb{N}, \tag{9.15}
\]
similar to (7.2), where \( \lambda_0 = 0.5 + \sqrt{3} \) as before. See Appendix A: Miscellaneous derivations.

In Figures 9.1 and 9.2 we compare the shifted c.d.f. \( F(w|\delta(0.5)) = I(w > 0.5)F_0(w - 0.5) \) involved for the case \( \gamma = 0.5 \) with \( F(w|\pi_n \delta(0.5)) \) for \( n = 20 \), on \([0, 1.5]\) and \([0, 0.5]\), respectively.

• Insert
  - Figure 9.1. Comparison of the shifted c.d.f. \( F(w|\delta(0.5)) \) (solid line) with \( F(w|\pi_n \delta(0.5)) \) (dotted line) for \( n = 20 \) on \([0, 1.5]\)
  - Figure 9.2. Comparison of the shifted c.d.f. \( F(w|\delta(0.5)) \) (solid line) with \( F(w|\pi_n \delta(0.5)) \) (dotted line) for \( n = 20 \) on \([0, 0.5]\)

Figure 9.1 suggests that \( F(w|\pi_n \delta(0.5)) \) approximates the left tail of \( F(w|\delta(0.5)) \) quite well, but zooming in on section \([0, 0.5]\) in Figure 9.2 we see that on \((0, 0.5],\)
0 < F(w|π_nδ(0.5)) ≤ 0.0062. The latter is very small, as expected, but not small enough to prevent \( F^{-1}(0|π_nδ(0.5)) = 0 \). On the other hand, in this case \( F^{-1}(0.0062|π_nδ(0.5)) = 0.5 \).

Although in finite samples the penalty function \( C.F^{-1}(0|π_nδ) \) may be ineffect, it is essential for the identification results in Theorem 3, and for the consistency of the sieve estimator of \( δ^0 \), as will be shown in the next section.

10. Integrated method of moments sieve estimation

10.1. Sieve estimation

First note that the number of solutions of \( \arg\min_{δ∈Δ} \hat{Q}_N(δ) \), with \( \hat{Q}_N(δ) \) defined by (9.8), is infinite because \( Δ \) is infinite dimensional, and that neither of these solutions will be consistent. The standard approach to avoid these problems is sieve estimation, proposed by Grenander (1981), as follows.

The idea of sieve estimation is to construct an increasing sequence of finite dimensional compact subsets \( Δ_n \), \( n ∈ N \), of \( Δ \), called sieve spaces, such that \( \bigcup_{n=1}^{∞} Δ_n = Δ \), where here the bar denotes the closure. For example, let

\[
Δ_n = \left\{ δ = \{δ_m\}_{m=1}^{∞} : \sum_{m=1}^{n} δ_m^2 ≤ M_n, \ δ_m = 0 \text{ for } m > n \right\}, \tag{10.1}
\]

where \( M_n \) is a given strictly increasing positive sequence converging to \( ∞ \) as \( n → ∞ \).

Next, let \( n_N ∈ N \) be any nondecreasing subsequence of the sample size \( N \) satisfying

\[
\lim_{N→∞} n_N = ∞, \ \lim_{N→∞} n_N/N = 0. \tag{10.2}
\]

Then

\[
\hat{δ}_N = \arg\min_{δ∈Δ_{n_N}} \hat{Q}_N(δ) \tag{10.3}
\]

is a sieve estimator of \( δ^0 \), 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 (10.2), the sieve estimator \( \hat{δ}_N \) is weakly consistent, in the sense that \( ||\hat{δ}_N - δ^0||_p \to 0 \) as \( N → ∞ \).
This result implies that $\hat{\alpha}_N = \overline{\alpha}_N(\delta_N) \overset{p}{\rightarrow} \alpha_0, \hat{\sigma}_N^2 = \overline{\sigma}_N^2(\delta_N) \overset{p}{\rightarrow} \sigma_0^2$, and

$$
\int_0^\infty |f(w;\delta_N) - f_0(w)|dw \overset{p}{\rightarrow} 0. \quad (10.4)
$$

10.2. Consistency of the sieve estimator

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), as will be shown now.

In general a crucial condition for the consistency of sieve estimators is that the objective function, $\overline{Q}_N(\delta) = \overline{Q}_N(\delta) + \Pi(\delta)$, where $\Pi(\delta) = C.F^{-1}(0|\delta)$, in our case, and its pointwise limit $Q(\delta) = \overline{Q}(\delta) + \Pi(\delta)$, are continuous in the parameter $\delta$. The continuity of $\overline{Q}_N(\delta)$ and $\overline{Q}(\delta)$ has already been established from Lemma 2, and the continuity of $\Pi(\delta) = F^{-1}(0|\delta)$ follows from Lemma 4. However, in the present case we only need the continuity of $Q(\delta)$ in $\delta_0$.

Our consistency proof is based on the following easy inequalities.

**Lemma 5.** For any $(\alpha, \sigma^2, \delta) \in \mathbb{R} \times \mathbb{R} \times \Delta$, and with $Q_N(\alpha, \sigma^2, \delta)$ and $Q(\alpha, \sigma^2, \delta)$ defined by $(8.10)$ and $(8.11)$, respectively, we have

$$
Q_N(\alpha, \sigma^2, \delta) \leq \sqrt{Q(\alpha, \sigma^2, \delta)} \left( \sqrt{Q(\alpha, \sigma^2, \delta)} + R_N \right), \quad (10.5)
$$

$$
Q_N(\alpha, \sigma^2, \delta) \geq \sqrt{Q(\alpha, \sigma^2, \delta)} \left( \sqrt{Q(\alpha, \sigma^2, \delta)} - R_N \right), \quad (10.6)
$$

where

$$
R_N = 2\sqrt{Q_N(\alpha_0, \sigma_0^2, \delta_0)^{a.s.}} \overset{a.s.}{\rightarrow} 0.
$$

Moreover,

$$
Q \left( \overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right) - Q \left( \overline{\alpha}(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right)
$$

$$
\leq S_N + 2\sqrt{S_N} \sqrt{Q \left( \overline{\alpha}(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right)}, \quad (10.7)
$$

$$
\geq S_N - 2\sqrt{S_N} \sqrt{Q \left( \overline{\alpha}(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right)}. \quad (10.8)
$$

\textsuperscript{24}C.f. (8.13) and (8.14).
where by (8.17),
\[ S_N = \int_0^1 \left( (\hat{\alpha}_N(\hat{\delta}_N) - \bar{\alpha}(\hat{\delta}_N)) t + (\hat{\sigma}_N^2(\hat{\delta}_N) - \bar{\sigma}^2(\hat{\delta}_N)) t^2 / 2 \right)^2 dt \xrightarrow{a.s.} 0. \]

**Proof.** See Appendix A. □

Denote for an arbitrary \( \varepsilon > 0 \),
\[ \Delta(\varepsilon) = \{ \delta \in \Delta : ||\delta - \delta^0|| \geq \varepsilon \} . \]

Then \( \hat{\delta}_N \in \Delta(\varepsilon) \) implies that
\[ Q(\hat{\delta}_N) \geq \inf_{\delta \in \Delta(\varepsilon)} Q(\delta) = \kappa > 0, \tag{10.9} \]
because \( \delta^0 \) is the only point for which \( Q(\delta^0) = 0 \).

Next, observe from inequality (10.6) that
\[
\hat{Q}_N(\hat{\delta}_N) = Q_N \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N)
\geq Q \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N) - R_N \sqrt{Q \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right)}
\geq \left( \sqrt{Q \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N) - R_N} \right)
\times \sqrt{Q \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N)}. \tag{10.10} \]

whereas by inequality (10.8) in Lemma 5 and (10.9)
\[
Q \left( \hat{\alpha}_N(\hat{\delta}_N), \hat{\sigma}_N^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N)
\geq Q \left( \hat{\alpha}(\hat{\delta}_N), \hat{\sigma}^2(\hat{\delta}_N), \hat{\delta}_N \right) + \Pi(\hat{\delta}_N)
\geq Q(\hat{\delta}_N) - 2 \sqrt{S_N} \sqrt{Q(\hat{\delta}_N)} = \left( \sqrt{Q(\hat{\delta}_N)} - 2 \sqrt{S_N} \right) \sqrt{Q(\hat{\delta}_N)}
\geq \left( \sqrt{\kappa} - 2 \sqrt{S_N} \right) \sqrt{\kappa} \text{ if } S_N < \kappa / 4. \tag{10.11} \]
Hence, it follows from (10.10) and (10.11) that
\[
\hat{Q}_N(\delta_N) \geq \left( \sqrt{Q\left(\alpha(\delta_N), \sigma^2(\delta_N), \delta_N\right)} + \Pi(\delta_N) - R_N \right) \\
\times \sqrt{Q\left(\alpha(\delta_N), \sigma^2(\delta_N), \delta_N\right)} + \Pi(\delta_N) \\
= \left( \sqrt{Q(\delta_N)} - R_N \right) \sqrt{Q(\delta_N)} \\
\geq \left( \sqrt{\kappa} \right) \left( \sqrt{Q(\delta_N)} \right) \\
\geq \left( \sqrt{\kappa - 2\sqrt{S_N}} \right) \left( \sqrt{\kappa - R_N} \right) \sqrt{\kappa} > 0
\]
if \( R_N < \sqrt{(\sqrt{\kappa - 2\sqrt{S_N}})\sqrt{\kappa}} \) and \( S_N < \kappa/4 \).

Note that \( \sqrt{\kappa - 2\sqrt{S_N}} \geq \kappa/2 \) if \( S_N \leq \kappa/16 \) and then
\[
\hat{Q}_N(\delta_N) \geq \sqrt{\kappa/2} - R_N \geq \kappa/4
\]
if \( R_N \leq \sqrt{\kappa/8} \). Since \( R_N \overset{a.s.}{\rightarrow} 0 \) and \( S_N \overset{a.s.}{\rightarrow} 0 \), there exists an \( N_0 \) such that \( S_N \leq \kappa/16 \) a.s. and \( R_N \leq \sqrt{\kappa/8} \) a.s. for all \( N \geq N_0 \). Thus, for \( N \geq N_0 \), \( \delta_N \in \Delta(\varepsilon) \) implies \( \hat{Q}_N(\delta_N) \geq \kappa/4 \), hence
\[
\Pr[\delta_N \in \Delta(\varepsilon)] \leq \Pr[\hat{Q}_N(\delta_N) \geq \kappa/4]. \tag{10.12}
\]

On the other hand, with \( N \) so large that \( \pi_{n_N} \delta^0 \in \Delta_{n_N} \), which is the case for \( M_{n_N} \geq ||\delta^0||^2 \) in (10.1), we have by inequality (10.5) that
\[
\hat{Q}_N(\delta_N) \leq \hat{Q}_N(\pi_{n_N} \delta^0) \\
= Q_N(\pi_{n_N} \delta^0, \sigma^2(\pi_{n_N} \delta^0), \pi_{n_N} \delta^0) + \Pi(\pi_{n_N} \delta^0) \\
\leq Q(\pi_{n_N} \delta^0, \sigma^2(\pi_{n_N} \delta^0), \pi_{n_N} \delta^0) + \Pi(\pi_{n_N} \delta^0) \\
+ \sqrt{Q(\pi_{n_N} \delta^0, \sigma^2(\pi_{n_N} \delta^0), \pi_{n_N} \delta^0)} \\
= Q(\pi_{n_N} \delta^0) + 2\sqrt{S_N} \sqrt{Q(\pi_{n_N} \delta^0)}.
\]

36
Thus,
\[
\hat{Q}_N(\delta_N) \leq Q(\pi_{nN}\delta_0) + 2\sqrt{S_N}\sqrt{Q(\pi_{nN}\delta_0)} + R_N. \\
\left(\sqrt{Q(\pi_{nN}\delta_0)} + 2\sqrt{S_N}\sqrt{Q(\pi_{nN}\delta_0)}\right)^{0.25} 0.
\]

Consequently, \(\lim_{N \to \infty} \Pr[\hat{Q}_N(\delta_N) \geq \kappa/4] = 0\), so that by (10.12), \(\lim_{N \to \infty} \Pr[\delta_N \in \Delta(\varepsilon)] = 0\), hence,
\[
\lim_{N \to \infty} \Pr[||\hat{\delta}_N - \delta_0|| \geq \varepsilon] = 0.
\]
Since \(\varepsilon > 0\) was arbitrary, the main result \(||\hat{\delta}_N - \delta_0|| \overset{p}{\to} 0\) follows. The results
\[
\hat{\alpha}_N = \alpha_N(\delta_N) \overset{p}{\to} \alpha_0 \quad \text{and} \quad \hat{\sigma}_N^2 = \sigma_N^2(\delta_N) \overset{p}{\to} \sigma_0^2
\]
follow easily from (8.17) and the continuity of \(\alpha(\delta)\) and \(\sigma^2(\delta)\) in \(\delta_0\). This completes the proof of Theorem 4.

**10.3. What to do if the penalty function is ineffective?**

As said before, the penalty function \(\Pi(\pi_{nN}\delta) = C.F^{-1}(0|\pi_{nN}\delta)\) may not be effective for finite \(N\). In that case the sieve estimator \(\hat{\delta}_N\) in (10.3) becomes
\[
\hat{\delta}_N = \arg \min_{\delta \in \Delta_{nN}} \overline{Q}_N(\delta),
\]
for which it is easy to show, along the lines in the previous subsection, that for an arbitrary \(\varepsilon > 0\), \(\lim_{N \to \infty} \Pr[\overline{Q}(\delta_N) \geq \varepsilon] = 0\), hence
\[
\overline{Q}(\delta_N) \overset{p}{\to} 0. \quad (10.13)
\]

At this point we could add the penalty function \(\Pi(\delta)\) to the objective function \(\overline{Q}_N(\delta)\) and minimize \(\hat{Q}_N(\delta) = \overline{Q}_N(\delta) + \Pi(\delta)\) further to \(\delta \in \Delta_{nN}\), starting from \(\delta = \hat{\delta}_N\), which yields a sieve estimator \(\hat{\delta}_N\) satisfying \(||\hat{\delta}_N - \delta_0|| \overset{p}{\to} 0\). This two-step sieve estimation procedure is equivalent to (10.3), but saves a substantial amount of computing time.

However, there is a much easier solution, which does not require a penalty function, via the following four steps.
1. First it will be shown below that for some $\gamma_N > 0$, possibly depending on $\delta_N$,

$$\int_0^\infty \left| f(w|\delta_N) - I(w > \gamma_N)f(w - \gamma_N|\delta^0) \right| dw = o_p(1).$$

$$\sup_{w>0} \left| F(w|\delta_N) - I(w > \gamma_N)F(w - \gamma_N|\delta^0) \right| = o_p(1).$$

2. Next, plot the c.d.f. $F(w|\delta_N)$, similar to Figures 9.1 and 9.2, and determine $\gamma_N$.\textsuperscript{25}

3. Then $\int_0^\infty \left| f(w + \gamma_N|\delta_N) - f(w|\delta^0) \right| dw = o_p(1)$, hence use $f(w + \gamma_N|\delta_N)$ as the consistent estimator of $f(w|\delta^0)$.

4. Finally, $\pi_N(\delta_N) - \gamma_N \overset{p}{\to} \alpha_0$ and $\sigma^2_N(\delta_N) \overset{p}{\to} \sigma^2_0$.

Strictly speaking, the consistency claims for $f(w + \gamma_N|\delta_N)$ and $\pi_N(\delta_N) - \gamma_N$ require that step 2 is repeated for each increase in $N$, because $\gamma_N$ is a random variable which varies with $N$. However, all asymptotic theories are based on planned actions in the case that the sample size $N$ increases, but in practice these actions are only taken once for the given sample, and the continuation of these actions for larger sample sizes is merely a thought experiment.

\textbf{10.3.1. Step 1}

Denote

$$\Delta_* = \{ \delta \in \Delta : \overline{\ell}(\delta) = 0 \},$$

and recall that every $\delta_* \in \Delta_*$ corresponds to an SNP density of the form

$$f(w|\delta_*) = I(w > \gamma)f_0(w - \gamma) = I(w > \gamma)f(w - \gamma|\delta^0)$$

for some $\gamma \geq 0$. Define the distance of any point $\delta \in \Delta$ to the set $\Delta_*$ by

$$d(\delta, \Delta_*) = \inf_{\delta_* \in \Delta_*} ||\delta - \delta_*||.$$

Clearly, $d(\delta, \Delta_*) = 0$ implies $\delta \in \Delta_*$, hence $\overline{\ell}(\delta) = 0$, and $d(\delta, \Delta_*) > 0$ implies $\delta \notin \Delta_*$, hence $\overline{\ell}(\delta) > 0$.

\textsuperscript{25}You may need a ruler to measure $\gamma_N$
Next, consider a deterministic sequence \( \delta_m \in \Delta, m \in \mathbb{N} \), such that
\[
\lim_{m \to \infty} Q(\delta_m) = 0.
\]
Then we must have \( \lim_{m \to \infty} d(\delta_m, \Delta_\ast) = 0 \). To see this, suppose that
\[
\lim \inf_{m \to \infty} d(\delta_m, \Delta_\ast) = \sup_{m \in \mathbb{N}} \inf_{k \geq m} d(\delta_k, \Delta_\ast) > \varepsilon > 0.
\]
This implies that there exists an \( m_0 \) such that for all \( m \geq m_0 \), \( d(\delta_m, \Delta_\ast) \geq \varepsilon \), hence \( Q(\delta_m) \geq \inf_{\delta \in \Delta_\ast} d(\delta, \Delta_\ast) \geq \varepsilon \) \( Q(\delta) > 0 \) for all \( m \geq m_0 \). However, the latter contradicts \( \lim_{m \to \infty} Q(\delta_m) = 0 \).

Similarly, (10.13) implies \( p \lim_{N \to \infty} d(\delta_N, \Delta_\ast) = 0 \). Moreover, denoting
\[
\delta_N^\ast = \arg \min_{\delta \in \Delta_\ast} ||\delta_N - \delta||
\]
so that \( d(\delta_N, \Delta_\ast) = ||\delta_N - \delta_N^\ast|| \), we now have
\[
p \lim_{N \to \infty} ||\delta_N - \delta_N^\ast|| = 0. \tag{10.14}
\]
Since \( \delta_N^\ast \in \Delta_\ast \), the corresponding SNP density \( f(w|\delta_N^\ast) \) and its c.d.f. \( F(w|\delta_N^\ast) \) take the form
\[
f(w|\delta_N^\ast) = I(w > \gamma_N) f_0(w - \gamma_N) = I(w > \gamma_N) f(w - \gamma_N|\delta^0),
\]
\[
F(w|\delta_N^\ast) = \int_0^w I(w > \gamma_N) F_0(w - \gamma_N) = I(w > \gamma_N) F(w - \gamma_N|\delta^0).
\]
where \( \gamma_N \) may depend on \( \delta_N^\ast \), whereas by Lemma 2 and (10.14),
\[
\int_0^\infty \left| f(w|\delta_N) - I(w > \gamma_N) f(w - \gamma_N|\delta^0) \right| dw \overset{p}{\to} 0,
\]
\[
\sup_{w > 0} \left| F(w|\delta_N) - I(w > \gamma_N) F(w - \gamma_N|\delta^0) \right| \overset{p}{\to} 0.
\]

10.3.2. Step 4

It follows from (9.3) and (9.4) that \( \alpha(\delta_N^\ast) - \gamma_N = \alpha_0 \) and \( \sigma^2(\delta_N^\ast) = \sigma_0^2 \), whereas similar to the argument in the footnote after (8.16), \( (\bar{\alpha}(\delta_N^\ast) - \bar{\alpha} (\delta_N)) \overset{p}{\to} 0 \) and \( (\bar{\sigma}^2(\delta_N^\ast) - \bar{\sigma}^2 (\delta_N)) \overset{p}{\to} 0 \), due to (10.14). Moreover, it follows from (8.17) that \( (\bar{\alpha}_N(\delta_N^\ast) - \bar{\alpha}(\delta_N^\ast)) \overset{p}{\to} 0 \) and \( (\bar{\sigma}_N^2(\delta_N^\ast) - \bar{\sigma}_N^2 (\delta_N)) \overset{p}{\to} 0 \). Combining these results, the results in Step 4 follow.
11. Prediction of inefficiency

For each firm \(i \in \mathbb{N}\) we have, under Assumption 1, and using the notations \(\hat{\alpha}_N = \overline{\alpha}_N(\hat{\delta}_N)\) and later on \(\hat{\sigma}_N^2 = \overline{\sigma}_N^2(\hat{\delta}_N)\),

\[
Y_i - \hat{\beta}_N X_i - \hat{\alpha}_N = a_0 - \hat{\alpha}_N - (\hat{\beta}_N - \beta_0)X_i + V_i - W_i
\]

\[
= V_i - W_i + o_p(1),
\]  
(11.1)
as is easy to verify. Thus, for each \(i \in \mathbb{N}\), \(\varepsilon_i = V_i - W_i\) can be estimated consistently by

\[
\hat{\varepsilon}_{N,i} = Y_i - \hat{\beta}_N X_i - \hat{\alpha}_N.
\]

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

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

where \(\phi\) is the standard normal density, hence the joint density of \(W\) and \(\varepsilon = V - W\) takes the form

\[
f_{W,\varepsilon}(w, \varepsilon) = \frac{1}{\sigma_0} \phi \left(\frac{\varepsilon + w}{\sigma_0}\right) f_0(w),
\]

with marginal densities \(f_W(w) = \int_{-\infty}^{\infty} f_{W,\varepsilon}(w, \varepsilon)d\varepsilon = f_0(w)\) and

\[
f_\varepsilon(\varepsilon) = \int_{0}^{\infty} f_{W,\varepsilon}(w, \varepsilon)dw = \int_{0}^{\infty} \frac{1}{\sigma_0} \phi \left(\frac{\varepsilon + w}{\sigma_0}\right) f_0(w)dw.
\]

Hence,

\[
E[\exp(-W_i)|V_i - W_i = \varepsilon] = \frac{\int_{0}^{\infty} \exp(-w) f_{W,\varepsilon}(w, \varepsilon)dw}{f_\varepsilon(\varepsilon)} = \eta_0(\varepsilon), \text{ say.}
\]

\(^{26}\)See for example Jondrow et al. (1982) and Horrace and Parmeter (2018).
Clearly, the TE function $\eta_0(\varepsilon)$ is continuous and uniformly bounded on $\mathbb{R}$.

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

$$
\hat{\eta}_N(\varepsilon) = \frac{\int_0^\infty \exp(-w)\phi((\varepsilon + w)/\hat{\sigma}_N)f(w|\delta_N)dw}{\int_0^\infty \phi((\varepsilon + w)/\hat{\sigma}_N)f(w|\delta_N)dw}
$$

(11.3)

as an estimate of the TE function $\eta_0(\varepsilon)$, for which the following results hold.

**Theorem 5.** Under Assumptions 1-2, $\sup_{|\varepsilon| \leq M} |\hat{\eta}_N(\varepsilon) - \eta_0(\varepsilon)| = o_p(1)$ for any $M \in (0, \infty)$ and $\hat{\eta}_N(\varepsilon_{N,i}) - \eta_0(\varepsilon_i) = o_p(1)$ for each $i \in \mathbb{N}$.

**Proof.** See Appendix A. □

Note that $\eta_0(\varepsilon_i) = E[\exp(-W_i)|\varepsilon_i] = E[\exp(-W_i)|V_i - W_i]$ is the actual TE index of firm $i$, so that the last result in Theorem 5 reads

$$
\hat{\eta}_N(Y_i - \hat{\beta}_N X_i - \hat{\alpha}_N) = E[\exp(-W_i)|V_i - W_i] + o_p(1).
$$

As to the computation of the integrals in (11.3), in view of the quadratic form (6.13) of $f(w|\delta_N)$, with $d_{k,m}(w)$ defined by (6.14), it suffices to compute the integrals

$$
\int_0^\infty \exp(-w)\phi\left(\frac{\varepsilon + w}{\hat{\sigma}_N}\right) d_{k,m}(w)dw
$$

$$
= \sum_{i=0}^k \sum_{j=0}^m \left(\begin{array}{c} k \\ i \end{array}\right) \left(\begin{array}{c} m \\ j \end{array}\right) \frac{(-1)^{i+j}}{i!j!} \int_0^\infty \phi\left(\frac{\varepsilon + w}{\hat{\sigma}_N}\right) w^{i+j} \exp(-2w)dw
$$

$$
= \sum_{i=0}^k \sum_{j=0}^m \left(\begin{array}{c} k \\ i \end{array}\right) \left(\begin{array}{c} m \\ j \end{array}\right) \frac{(-1)^{i+j}}{i!j!} 2^{-i-j-1} \int_0^\infty \phi\left(\frac{\varepsilon + w/2}{\hat{\sigma}_N}\right) w^{i+j} \exp(-w)dw
$$

$$
= \sum_{i=0}^k \sum_{j=0}^m \left(\begin{array}{c} k \\ i \end{array}\right) \left(\begin{array}{c} m \\ j \end{array}\right) \frac{(-1)^{i+j}}{i!j!} 2^{-i-j-1} \times \int_0^\infty \phi\left(\frac{\varepsilon + w/2}{\hat{\sigma}_N}\right) w^{i+j} \exp(-w)dw
$$

and similarly,

$$
\int_0^\infty \phi\left(\frac{\varepsilon + w}{\hat{\sigma}_N}\right) d_{k,m}(w)dw
$$

41
\[ \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} (i + j)^{-1} \int_{0}^{\infty} \phi \left( \frac{\varepsilon + w}{\sigma_N} \right) \frac{w^{i+j}}{(i+j)!} \exp(-w) dw \]

for \( k, m = 0, 1, 2, ..., n_N \), where each of the integrals

\[ \int_{0}^{\infty} \phi \left( \frac{\varepsilon + w/2}{\sigma_N} \right) \frac{w^{\ell}}{\ell!} \exp(-w) dw \]

\[ = \int_{0}^{\infty} \phi \left( \frac{\varepsilon + w/2}{\sigma_N} \right) \frac{w^{\ell}}{\ell!} d(1 - \exp(-w)) \]

\[ = \int_{0}^{1} \phi \left( \frac{\varepsilon + \frac{1}{2} \ln(1/(1-u))}{\sigma_N} \right) \frac{(\ln(1/(1-u)))^{\ell}}{\ell!} du \]  
(11.4)

and

\[ \int_{0}^{\infty} \phi \left( \frac{\varepsilon + w}{\sigma_N} \right) \frac{w^{\ell}}{\ell!} \exp(-w) dw \]

\[ = \int_{0}^{1} \phi \left( \frac{\varepsilon + \ln(1/(1-u))}{\sigma_N} \right) \frac{(\ln(1/(1-u)))^{\ell}}{\ell!} du \]  
(11.5)

for \( \ell = 0, 1, ..., 2n_N \) can be computed by the LGQ approximation. See Appendix B for the latter.

Finally, note that

\[ \frac{w^{\ell}}{\ell!} \exp(-w) < \sum_{m=0}^{\infty} \frac{w^{m}}{m!} \exp(-w) = 1, \]

so that the integrands in (11.4) and (11.5) are bounded. This is advantageous for the application of the LGQ approximations of (11.4) and (11.5).

12. Application to the numerical example

12.1. To be done!

13. An empirical application

In this section, we (plan to) 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), Hoffer 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 group. The dependent variable is the logarithm of monthly wage ($Y$), and the independent variables include the human capital variables: years of base schooling ($Edu_1$), years of higher education ($Edu_2$), experience ($Exp$), and the squared experience ($Exp^2$). 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 11.1 below.

The model is specified as

$$Y^j = \alpha^j_0 + \beta^j_1 Edu_1 + \beta^j_2 Edu_2 + \beta^j_3 Exp + \beta^j_4 Exp^2 + \beta^j_5 Scale + V^j - W^j$$

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 11.1. Definitions of the variables for the Taiwan labor market data**

<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>$Edu_1$</td>
<td>Years of base-education schooling (including elementary, junior and senior high schools)</td>
</tr>
<tr>
<td>$Edu_2$</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 $Exp = \min{\text{Age-15}, \text{Age-Years of Education-8}}$</td>
</tr>
<tr>
<td>$Exp^2$</td>
<td>Squared experience divided by 100, i.e., $(Exp \times Exp)/100$</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>
13.1. To be continued!

14. Concluding remarks

14.1. To be done!
References


Fan, Y., Q. Li, and A. Weersink (1996), "Semiparametric estimation of sto-

\(^{27}\)Downloadable from http://www.personal.psu.edu/hxb11/HANDBOOK CHAPTER.PDF


Heckman, J. J. (1979), ”Sample selection bias as a specification error,” Econometrica 47, 153-161.


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


15. Appendix A: Proofs

15.1. Theorem 2

Since the result (4.8) in Theorem 1 follows straightforwardly from Lemma 1, we only need to prove the results (4.9) and (4.10), as follows.

It follows trivially from (4.7) that

\[ \gamma'_N(t) = -\sum_{i=1}^{N} (Z_{i,N} + |X_i|) \exp(-t.(Z_{i,N} + |X_i|)) + \frac{\sum_{i=1}^{N} |X_i| \exp(-t.|X_i|)}{\sum_{i=1}^{N} \exp(-t.|X_i|)} \]

where \( Z_{i,N} = X_i^{\prime}(\beta_{N} - \beta_0) + W_i - V_i - \alpha_0 \).

By a similar argument as in the proof of Lemma 1 it follows that

\[ \sup_{0 \leq t \leq c} \left| \frac{1}{N} \sum_{i=1}^{N} (Z_{i,N} + |X_i|) \exp(-t.(Z_{i,N} + |X_i|)) \right| \to 0 \]

whereas by Jennrich’s (1969) uniform strong law of large numbers,

\[ \frac{1}{N} \sum_{i=1}^{N} (W_i - V_i - \alpha_0 + |X_i|) \exp(-t.(W_i - V_i - \alpha_0 + |X_i|)) \to E \left[ (W - V - \alpha_0 + |X|) \exp(-t.(W - V - \alpha_0 + |X|)) \right] \]

uniformly on [0, 1]. Due to the independence of \( W, V \) and \(|X|\) the latter expectation can be written as

\[ E \left[ (W - V - \alpha_0 + |X|) \exp(-t.(W - V - \alpha_0 + |X|)) \right] = E[W \exp(-t.W) \exp(t.V) \exp(t_0) \exp(-t.|X|)] - E[V \exp(t.V) \exp(-t.W) \exp(t_0) \exp(-t.|X|)] - \alpha_0 \exp(t_0) \exp(-t.W) \exp(t.V) \exp(-t.|X|)] + E[|X| \exp(-t.|X|)] \exp(t_0) \exp(-t.W) \exp(t.V)

= E[W \exp(-t.W) \exp(\sigma^2 t^2/2) \exp(t_0) \exp(-t.|X|)] - t \sigma^2 \exp(\sigma^2 t^2/2) \exp(t_0) \exp(-t.W) \exp(-t.|X|)] - \alpha_0 \exp(t_0) \exp(-t.W) \exp(\sigma^2 t^2/2) \exp(-t.|X|)] + E[|X| \exp(-t.|X|)] \exp(t_0) \exp(-t.W) \exp(\sigma^2 t^2/2) \]
where the last equality is due to $E[\exp(t.V)] = \exp(\sigma^2_0 t^2 / 2)$, hence $E[V \exp(t.V)] = d\exp(\sigma^2_0 t^2 / 2)/dt = t.\sigma^2_0 \exp(\sigma^2_0 t^2 / 2)$.

It follows now from Lemma 1 that

$$\sum_{i=1}^N (Z_{i,N} + \|X_i\|) \exp(-t.(Z_{i,N} + \|X_i\|)) \over \sum_{i=1}^N \exp(-t.(Z_{i,N} + \|X_i\|)) \xrightarrow{a.s.} E[W \exp(-t.W)] + \frac{E[\|X\| \exp(-t.|X|)]}{E[\exp(-t.|X|)]} - \alpha_0 - t.\sigma^2_0$$

uniformly on $[0, 1]$. Since also

$$\sum_{i=1}^N \|X_i\| \exp(-t.|X_i|) \over \sum_{i=1}^N \exp(-t.|X_i|) \xrightarrow{a.s.} \frac{E[\|X\| \exp(-t.|X|)]}{E[\exp(-t.|X|)]}$$

uniformly on $[0, 1]$, the result (4.9) follows.

Along similar lines it can be shown that

$$\sup_{0 \leq t \leq c} \left| \sum_{i=1}^N \|X_i\|^2 \exp(-t.|X_i|) \over \sum_{i=1}^N \exp(-t.|X_i|) - \frac{E[\|X\|^2 \exp(-t.|X|)]}{E[\exp(-t.|X|)]} \right| \xrightarrow{a.s.} 0$$

and

$$\sup_{0 \leq t \leq c} \left| \sum_{i=1}^N (Z_{i,N} + \|X_i\|)^2 \exp(-t.(Z_{i,N} + \|X_i\|)) \over \sum_{i=1}^N \exp(-t.(Z_{i,N} + \|X_i\|)) - \left( \frac{\sum_{i=1}^N (Z_{i,N} + \|X_i\|) \exp(-t.(Z_{i,N} + \|X_i\|))}{\sum_{i=1}^N \exp(-t.(Z_{i,N} + \|X_i\|))} \right)^2 \right| \xrightarrow{a.s.} 0,$$

which after some tedious but elementary calculations delivers the result (4.10).
15.2. Lemma 2

Observe from (6.7) that for every pair \( \mathbf{\delta}_1 = \{\delta_{1,m}\}_{m=1}^{\infty}, \mathbf{\delta}_2 = \{\delta_{2,m}\}_{m=1}^{\infty} \) in \( \Delta \),

\[
\begin{align*}
f(w|\mathbf{\delta}_1) - f(w|\mathbf{\delta}_2) \\
= \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right)^2 \\
- \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right)^2 \\
= \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right) - \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right)^2 \\
+ \exp(-w) \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2} \right)^2 - \left( \frac{1 + \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right)^2 \\
= \left( 1 - \frac{1 + \sum_{m=1}^{\infty} \delta_{1,m}^2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\mathbf{\delta}) \\
+ \frac{1}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
\times \left( 2 + \sum_{m=1}^{\infty} (\delta_{1,m} + \delta_{2,m}) \rho_m(w) \right) \\
= \left( \frac{\sum_{m=1}^{\infty} (\delta_{2,m}^2 - \delta_{1,m}^2)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\mathbf{\delta}) \\
+ \frac{1}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
\times \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) + 2 \sum_{m=1}^{\infty} \delta_{2,m} \rho_m(w) \right) \\
+ \frac{2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_m(w) \right) \\
= \left( \frac{\sum_{m=1}^{\infty} (\delta_{2,m}^2 - \delta_{1,m}^2)}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^2} \right) f(w|\mathbf{\delta})
\end{align*}
\]
\[ + \frac{1}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^{2}} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_{m}(w) \right)^{2} \]

\[ + \frac{2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^{2}} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_{m}(w) \right) \times \left( \sum_{m=1}^{\infty} \delta_{2,m} \rho_{m}(w) \right) \]

\[ + \frac{2}{1 + \sum_{m=1}^{\infty} \delta_{2,m}^{2}} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_{m}(w) \right). \]

Next, observe that

\[ \sum_{m=1}^{\infty} (\delta_{2,m}^{2} - \delta_{1,m}^{2}) = \sum_{m=1}^{\infty} (\delta_{2,m}^{2} - \delta_{1,m}) (\delta_{2,m}^{2} + \delta_{1,m}) \]

\[ \leq \sum_{m=1}^{\infty} (\delta_{2,m}^{2} - \delta_{1,m})^{2} + 2 \sum_{m=1}^{\infty} (\delta_{2,m}^{2} - \delta_{1,m}) \delta_{2,m}^{2} \]

\[ = \| \delta_{2} - \delta_{1} \|^{2} + 2|\langle \delta_{2} - \delta_{1}, \delta_{2} \rangle| \]

\[ \leq \| \delta_{2} - \delta_{1} \|^{2} + 2\| \delta_{2} - \delta_{1} \| \| \delta_{2} \| \]

where the second equality follows from (6.9) and (6.10), 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

\[ \int_{0}^{\infty} |f(w|\delta_{1}) - f(w|\delta_{2})| \, dw \]

\[ \leq \frac{\| \delta_{2} - \delta_{1} \|^{2} + 2\| \delta_{2} - \delta_{1} \| \| \delta_{2} \|}{1 + \| \delta_{2} \|^{2}} \]

\[ + \frac{1}{1 + \| \delta_{2} \|^{2}} \int_{0}^{\infty} \exp(-w) \left( \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_{m}(w) \right)^{2} \, dw \]

\[ + \frac{2}{1 + \| \delta_{2} \|^{2}} \int_{0}^{\infty} \exp(-w) \left| \sum_{m=1}^{\infty} (\delta_{1,m} - \delta_{2,m}) \rho_{m}(w) \right| \times \sum_{m=1}^{\infty} \delta_{2,m} \rho_{m}(w) \, dw \]

51
+ \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||\cdot||\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\sqrt{\int_0^\infty \exp(-w) \left( \sum_{m=1}^\infty (\delta_{1,m} - \delta_{2,m})\rho_m(w) \right)^2 \, dw}
\times \sqrt{\int_0^\infty \exp(-w) \left( \sum_{m=1}^\infty \delta_{2,m}\rho_m(w) \right)^2 \, dw}
+2\sqrt{\int_0^\infty \exp(-w) \left( \sum_{m=1}^\infty (\delta_{1,m} - \delta_{2,m})\rho_m(w) \right)^2 \, dw}
\leq \frac{1}{1 + ||\delta_2||^2} \left( ||\delta_2 - \delta_1||^2 + 2||\delta_2 - \delta_1||\cdot||\delta_2|| + 2||\delta_2 - \delta_1|| \right)
\leq 2||\delta_2 - \delta_1||^2 + 4||\delta_2 - \delta_1|| \cdot \frac{||\delta_2||}{1 + ||\delta_2||^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 \( d(x(1 + x^2)^{-1})/dx = 0 \) for \( x = 1 \), and \( d^2(x(1 + x^2)^{-1})/(dx)^2 \big|_{x=1} = -6 \), hence

\[
\frac{||\delta_2||}{1 + ||\delta_2||^2} \leq \max_{x \geq 0} \frac{x}{1 + x^2} = \frac{1}{2}.
\]

**15.3. Lemma 3**

Let \( n \) be so large that \( ||\pi_n \delta - \delta|| < 1 \), so that by Lemma 2,

\[
\int_0^\infty |f(w|\pi_n \delta) - f(w|\delta)| \, dw \leq 6.||\pi_n \delta - \delta||,
\]
and let

$$K_n = ||\pi_n\delta - \delta||^{-1/(2\ell)}.$$  

Then

$$\left| \int_0^{K_n} w^\ell f(w|\pi_n\delta)dw - \int_0^{K_n} w^\ell f(w|\delta)dw \right| \leq K_n^\ell \int_0^\infty |f(w|\pi_n\delta) - f(w|\delta)|dw \leq 6\sqrt{||\pi_n\delta - \delta||}. \quad (15.1)$$

This inequality carries over for any $M > 0$ and $n$ so large that that $K_n > M$ and $||\pi_n\delta - \delta|| < 1$, as

$$\left| \int_0^M w^\ell f(w|\pi_n\delta)dw - \int_0^M w^\ell f(w|\delta)dw \right| \leq 6\sqrt{||\pi_n\delta - \delta||}. \quad (15.2)$$

Suppose in first instance that $\int_0^\infty w^\ell f(w|\delta)dw = \infty$, so that

$$\lim_{n\to\infty} \int_0^{K_n} w^\ell f(w|\delta)dw = \infty.$$  

Then by (15.1),

$$\int_0^\infty w^\ell f(w|\pi_n\delta)dw \geq \int_0^{K_n} w^\ell f(w|\pi_n\delta)dw \geq \int_0^{K_n} w^\ell f(w|\delta)dw - 6\sqrt{||\pi_n\delta - \delta||} \to \infty \text{ as } n \to \infty.$$  

Next, suppose that $\int_0^\infty w^\ell f(w|\delta)dw < \infty$, which implies

$$\lim_{M\to\infty} \int_0^M w^\ell f(w|\delta)dw = \int_0^\infty w^\ell f(w|\delta)dw$$

Then by (15.2), for $n$ so large that that $K_n > M$ and $||\pi_n\delta - \delta|| < 1$

$$\int_0^M w^\ell f(w|\pi_n\delta)dw \leq \int_0^M w^\ell f(w|\delta)dw + 6\sqrt{||\pi_n\delta - \delta||},$$

hence

$$\int_0^M \inf_{k\geq n} w^\ell f(w|\pi_k\delta)dw \leq \int_0^M w^\ell f(w|\pi_n\delta)dw \leq \int_0^M w^\ell f(w|\delta)dw + 6\sqrt{||\pi_n\delta - \delta||}. \quad (15.3)$$
Since the sequence \( \inf_{k \geq n} f(w|\pi_k \delta) \) is monotonic non-decreasing, pointwise in \( w \), it follows from the monotonic convergence theorem\(^{28}\) that
\[
\lim_{n \to \infty} \int_0^M w^\ell \inf_{k \geq n} f(w|\pi_k \delta) \, dw = \int_0^M w^\ell \lim_{n \to \infty} \inf_{k \geq n} f(w|\pi_k \delta) \, dw. \tag{15.4}
\]
Moreover, it follows from Theorem 16 and its Corollary 2 in Bierens (2014a) that
\[
\lim_{n \to \infty} f(w|\pi_n \delta) = f(w|\delta) \text{ a.e. on } [0, \infty),
\]
so that
\[
\lim_{n \to \infty} \inf_{k \geq n} f(w|\pi_k \delta) = \lim_{n \to \infty} f(w|\pi_n \delta) = f(w|\delta). \tag{15.5}
\]
Thus, (15.3), (15.4) and (15.5) imply that for all \( M > 0 \),
\[
\int_0^M w^\ell f(w|\delta) \, dw \leq \lim_{n \to \infty} \int_0^M w^\ell f(w|\pi_n \delta) \, dw \leq \int_0^M w^\ell f(w|\delta) \, dw.
\]
Letting \( M \to \infty \), it follows that
\[
\lim_{n \to \infty} \int_0^\infty w^\ell f(w|\pi_n \delta) \, dw = \int_0^\infty w^\ell f(w|\delta) \, dw.
\]

15.4. Lemma 4

Let \( F(w|\delta) \), \( w \geq 0 \), be an absolutely continuous SNP distribution function, where \( \delta \in \Delta \). For \( u \in [0, 1) \) and \( \delta \in \Delta \), denote \( w(u|\delta) = F^{-1}(u|\delta) \). Note that by the definition (9.5) of \( F^{-1}(u|\delta) \), \( w(u|\delta) \) is strictly monotonic increasing in \( u \in [0, 1) \).

The question is: how is \( |w(u|\delta) - w(u|\delta_*)| \) related to \( ||\delta - \delta_*|| \)?

Recall from Lemma 2 that
\[
\sup_{w > 0} |F(w|\delta) - F(w|\delta_*)| \leq \int_0^\infty |f(w|\delta) - f(w|\delta_*)| \, dw \leq 2||\delta - \delta_*||^2 + 4||\delta - \delta_*|| < 6||\delta - \delta_*|| \tag{15.6}
\]
if \( ||\delta - \delta_*|| < 1 \). The latter will be assumed.

Suppose that for some \( u \in [0, 1) \) and an arbitrarily small \( \epsilon > 0 \), \( |w(u|\delta_*) - w(u|\delta)| \geq \epsilon \), so that either
\[
w(u|\delta_*) \geq w(u|\delta) + \epsilon \tag{15.7}
\]
\(^{28}\)See for example, Bierens (2004, Theorem 2.10).
or

$$w(u|\delta) \leq w(u|\delta^*) - \varepsilon.$$ \hfill (15.8)

In the case (15.7) we have

$$u = F(w(u|\delta^*)|\delta^*)$$

$$\geq F(w(u|\delta) + \varepsilon|\delta^*)$$

$$= F(w(u|\delta) + \varepsilon|\delta^*) - F(w(u|\delta) + \varepsilon|\delta) + F(w(u|\delta) + \varepsilon|\delta)$$

hence by (15.6) and $$||\delta - \delta^*|| \to 0,$$

$$u - F(w(u|\delta) + \varepsilon|\delta) \geq F(w(u|\delta) + \varepsilon|\delta^*) - F(w(u|\delta) + \varepsilon|\delta)$$

$$> -6||\delta - \delta^*|| \to 0.$$ 

Thus,

$$F(w(u|\delta) + \varepsilon|\delta) \leq u.$$ \hfill (15.9)

However, this inequality is impossible, even for $$u = 0$$. To see this, note that (15.9) together with the trivial inequality $$F(w(u|\delta) + \varepsilon|\delta) \geq F(w(u|\delta)|\delta) = u$$ imply $$F(w(u|\delta) + \varepsilon|\delta) = u = F(w(u|\delta)|\delta),$$ hence, $$F(w|\delta) = u$$ for all $$w \in (w(u|\delta), w(u|\delta) + \varepsilon)$$ and thus

$$w(u|\delta) \overset{\text{def.}}{=} \inf_{w > 0 : F(w|\delta) > u} w \geq w(u|\delta) + \varepsilon.$$ 

This contradiction implies that for each $$\delta \in \Delta$$ and each $$u \in [0, 1),$$ $$w(u|\delta^*) - w(u|\delta) < \varepsilon$$ as $$||\delta - \delta^*|| \to 0.$$

Similarly, in the case (15.8), $$u - F(w(u|\delta) - \varepsilon|\delta) \leq 6||\delta - \delta^*|| \to 0,$$ hence

$$F(w(u|\delta) - \varepsilon|\delta) \geq u,$$ \hfill (15.10)

whereas $$F(w(u|\delta) - \varepsilon|\delta) \leq F(w(u|\delta)|\delta) = u.$$ Hence, a similar argument as before yields the contradiction

$$w(u|\delta) \overset{\text{def.}}{=} \inf_{w > 0 : F(w|\delta) > u} w \leq w(u|\delta) - \varepsilon.$$ 

Thus, for each $$\delta \in \Delta$$ and each $$u \in [0, 1),$$ $$w(u|\delta^*) - w(u|\delta) > -\varepsilon$$ as $$||\delta - \delta^*|| \to 0.$$

Consequently, for each $$\delta \in \Delta$$ and each $$u \in [0, 1),$$ $$w(u|\delta^*) \to w(u|\delta) \to 0$$ as $$||\delta - \delta^*|| \to 0.$$
15.5. Lemma 5

Recall from (8.10) and (8.11) that

\[ Q_N(\alpha, \sigma^2, \delta) = \int_0^1 \psi_N(t|\alpha, \sigma^2, \delta)^2 dt, \]

where

\[ \psi_N(t|\alpha, \sigma^2, \delta) = \Upsilon_N(t) - \alpha t - \sigma^2 t^2/2 - \ln(\mathcal{L}(t|\delta)), \]

and

\[ Q(\alpha, \sigma^2, \delta) = \int_0^1 \psi(t|\alpha, \sigma^2, \delta)^2 dt, \]

where

\[ \psi(t|\alpha, \sigma^2, \delta) = \ln(\mathcal{L}(t|\delta^0)) - \ln(\mathcal{L}(t|\delta)) - (\alpha - \alpha_0) t - (\sigma^2 - \sigma_0^2) t^2/2, \]

so that trivially,

\[ \psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta) = \psi_N(t|\alpha_0, \sigma_0^2, \delta^0). \]

Then

\[
Q_N(\alpha, \sigma^2, \delta) - Q(\alpha, \sigma^2, \delta) \\
= \int_0^1 (\psi_N(t|\alpha, \sigma^2, \delta) - \psi(t|\alpha, \sigma^2, \delta)) \\
\times (\psi_N(t|\alpha, \sigma^2, \delta) + \psi(t|\alpha, \sigma^2, \delta)) dt \\
= \int_0^1 \psi_N(t|\alpha_0, \sigma_0^2, \delta^0) \\
\times (\psi_N(t|\alpha_0, \sigma_0^2, \delta^0) + 2\psi(t|\alpha, \sigma^2, \delta)) dt \\
= Q_N(\alpha_0, \sigma_0^2, \delta^0) + 2 \int_0^1 \psi_N(t|\alpha_0, \sigma_0^2, \delta^0) \psi(t|\alpha, \sigma^2, \delta) dt. \tag{15.11}
\]

Since by the Cauchy-Schwarz inequality,

\[
\left| \int_0^1 \psi_N(t|\alpha_0, \sigma_0^2, \delta^0) \psi(t|\alpha, \sigma^2, \delta) dt \right| \\
\leq \sqrt{\int_0^1 \psi_N(t|\alpha_0, \sigma_0^2, \delta^0)^2 dt} \sqrt{\int_0^1 \psi(t|\alpha, \sigma^2, \delta)^2 dt} \\
= \sqrt{Q_N(\alpha_0, \sigma_0^2, \delta^0)} \sqrt{Q(\alpha, \sigma^2, \delta)}
\]

the inequalities (10.5) and (10.6) in Lemma 5 follow.
The inequalities (10.7) and (10.8) follow similar to (15.11), i.e.,

\[ Q \left( \overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right) - Q \left( \overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N \right) \]

\[ = \int_0^1 \left( \psi(t|\overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N) \right)^2 dt \]

\[ - \int_0^1 \left( \psi(t|\overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N) \right)^2 dt \]

\[ = \int_0^1 \left( \psi(t|\overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N) - \psi(t|\overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N) \right)^2 dt \]

\[ + 2 \int_0^1 \left( \psi(t|\overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N) - \psi(t|\overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N) \right) \times \psi(t|\overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N) dt \]

\[ = \int_0^1 \left( \left( \overline{\alpha}(\delta_N) - \overline{\alpha}_N(\delta_N) \right) t + \left( \overline{\sigma}^2(\delta_N) - \overline{\sigma}_N^2(\delta_N) \right) t^2/2 \right)^2 dt \]

\[ + 2 \int_0^1 \left( \left( \overline{\alpha}(\delta_N) - \overline{\alpha}_N(\delta_N) \right) t + \left( \overline{\sigma}^2(\delta_N) - \overline{\sigma}_N^2(\delta_N) \right) t^2/2 \right) \times \psi(t|\overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N) dt \]

\[ \leq S_N + 2\sqrt{S_N} \sqrt{Q \left( \overline{\alpha}(\delta_N), \overline{\sigma}^2(\delta_N), \hat{\delta}_N \right)}, \]

\[ \geq S_N - 2\sqrt{S_N} \sqrt{Q \left( \overline{\alpha}_N(\delta_N), \overline{\sigma}_N^2(\delta_N), \hat{\delta}_N \right)} \]

15.6. Theorem 5

It will first be shown that for \( \varepsilon \in \mathbb{R} \),

\[ \sup_{\varepsilon \in \mathbb{R}} \left| \hat{f}_{N, \varepsilon}(\varepsilon) - f_{\varepsilon}(\varepsilon) \right| = o_p(1), \quad (15.12) \]

as follows. Observe that

\[ \left| \hat{f}_{N, \varepsilon}(\varepsilon) - f_{\varepsilon}(\varepsilon) \right| \]

\[ = \left| \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{\varepsilon + w}{\sigma_N} \right) f(w|\hat{\delta}_N) dw - \int_0^\infty \frac{1}{\sigma_0} \phi \left( \frac{\varepsilon + w}{\sigma_0} \right) f_0(w) dw \right| \]

\[ \leq \int_0^\infty \frac{1}{\sigma_N} \phi \left( \frac{\varepsilon + w}{\sigma_N} \right) \left| f(w|\hat{\delta}_N) - f_0(w) \right| dw \]
\[
+ \int_0^\infty \left| \frac{1}{\hat{\sigma}_N} \phi \left( \frac{\varepsilon + w}{\hat{\sigma}_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{\varepsilon + w}{\sigma_0} \right) \right| f_0(w)dw \\
\leq \frac{1}{\hat{\sigma}_N \sqrt{2\pi}} \int_0^\infty \left| f(w) - f_0(w) \right| dw \\
+ \int_0^\infty \left| \frac{1}{\sigma_N} \phi \left( \frac{\varepsilon + w}{\sigma_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{\varepsilon + w}{\sigma_0} \right) \right| f_0(w)dw. \tag{15.13}
\]

If \( \hat{\sigma}_N < \sigma_0 \) then
\[
\left| \frac{1}{\hat{\sigma}_N} \phi \left( \frac{\varepsilon + w}{\hat{\sigma}_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{\varepsilon + w}{\sigma_0} \right) \right| \\
\leq \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\hat{\sigma}_N^2 \right)}{\hat{\sigma}_N \sqrt{2\pi}} - \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\sigma_0^2 \right)}{\hat{\sigma}_N \sqrt{2\pi}} \\
+ \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\sigma_0^2 \right)}{\hat{\sigma}_N \sqrt{2\pi}} - \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\sigma_0^2 \right)}{\sigma_0 \sqrt{2\pi}} \\
\leq \exp \left( -\frac{1}{2} (\varepsilon + w)^2 \left( \hat{\sigma}_N^{-2} - \sigma_0^{-2} \right) \right) - 1 \left| \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\sigma_0^2 \right)}{\hat{\sigma}_N \sqrt{2\pi}} \right| \\
+ \frac{1}{\sqrt{2\pi}} \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right| \\
\leq \frac{|\hat{\sigma}_N^{-2} - \sigma_0^{-2}| \sigma_0^2}{\hat{\sigma}_N \sqrt{2\pi}} \exp \left( -\frac{1}{2} (\varepsilon + w)^2/\sigma_0^2 \right) + \frac{1}{\sqrt{2\pi}} \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right| \\
\leq \frac{|\hat{\sigma}_N^{-2} - \sigma_0^{-2}| \sigma_0^2}{\hat{\sigma}_N \sqrt{2\pi}} \exp(-1) + \frac{1}{\sqrt{2\pi}} \left| \hat{\sigma}_N^{-1} - \sigma_0^{-1} \right|, \tag{15.14}
\]

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
\[
\left| \frac{1}{\hat{\sigma}_N} \phi \left( \frac{\varepsilon + w}{\hat{\sigma}_N} \right) - \frac{1}{\sigma_0} \phi \left( \frac{\varepsilon + w}{\sigma_0} \right) \right| \\
\leq \exp \left( -\frac{1}{2} (\varepsilon + w)^2/\hat{\sigma}_N^2 \right) - \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\hat{\sigma}_N^2 \right)}{\sigma_0 \sqrt{2\pi}} \\
\leq \frac{\exp \left( -\frac{1}{2} (\varepsilon + w)^2/\hat{\sigma}_N^2 \right)}{\sigma_0 \sqrt{2\pi}}.
\]
It follows now straightforwardly from (15.13), (15.14), (15.15) and Theorem 3 but hence, (15.16) that

\[ \sup_{\varepsilon} \left| f_{N,\varepsilon}(\varepsilon) \tilde{\eta}_N(\varepsilon) - f_{\varepsilon}(\varepsilon) \eta_0(\varepsilon) \right| = \sup_{\varepsilon} \left| \int_0^\infty \exp(-w) \hat{f}_{N,\varepsilon}(w,\varepsilon) \, dw \right| - \left| \int_0^\infty \exp(-w) f_{W,\varepsilon}(w,\varepsilon) \, dw \right| = o_p(1), \]

hence,

\[ f_{\varepsilon}(\varepsilon) \cdot \tilde{\eta}_N(\varepsilon) = f_{\varepsilon}(\varepsilon) \eta_0(\varepsilon) \]

\[ = \left| f_{N,\varepsilon}(\varepsilon) \tilde{\eta}_N(\varepsilon) - f_{\varepsilon}(\varepsilon) \eta_0(\varepsilon) \right| - \left( f_{N,\varepsilon}(\varepsilon) - f_{\varepsilon}(\varepsilon) \right) \cdot \tilde{\eta}_N(\varepsilon) \]

\[ \leq \sup_{\varepsilon} \left| f_{N,\varepsilon}(\varepsilon) \tilde{\eta}_N(\varepsilon) - f_{\varepsilon}(\varepsilon) \eta_0(\varepsilon) \right| + \sup_{\varepsilon} \left| \tilde{\eta}_N(\varepsilon) \right| = o_p(1), \quad (15.16) \]

where the inequality is due to \( \sup_{\varepsilon} \tilde{\eta}_N(\varepsilon) \leq 1. \)

Now observe from (11.2) that by bounded convergence, \( \lim_{|\varepsilon| \to \infty} f_{\varepsilon}(\varepsilon) = 0, \)

but \( f_{\varepsilon}(\varepsilon) > 0 \) for all \( \varepsilon \in \mathbb{R}. \) Since \( f_{\varepsilon}(\varepsilon) \) is continuous on \( \mathbb{R} \) it follows therefore that for any constant \( M > 0, \inf_{|\varepsilon| \leq M} f_{\varepsilon}(\varepsilon) > 0, \) which in its turn implies by (15.16) that

\[ \sup_{|\varepsilon| \leq M} \left| \tilde{\eta}_N(\varepsilon) - \eta_0(\varepsilon) \right| \leq \frac{\sup_{|\varepsilon| \leq M} f_{\varepsilon}(\varepsilon) \cdot \tilde{\eta}_N(\varepsilon) - \eta_0(\varepsilon)}{\inf_{|\varepsilon| \leq M} f_{\varepsilon}(\varepsilon)} = o_p(1). \quad (15.17) \]

59
As has been shown before, for each $i$, $\varepsilon_i = V_i - W_i$ can be estimated consistently by $\hat{\varepsilon}_{N,i} = Y_i - \hat{\beta}_N X_i - \hat{\alpha}_N$. It will now be shown that this implies that $\hat{\eta}(\hat{\varepsilon}_{N,i}) - \eta_0(\varepsilon_i) = o_p(1)$, as follows. Let $\kappa \in (0, 1)$ be arbitrary, and let $M > 0$ be so large that $\Pr(|\varepsilon_i| > M/2) < \kappa/2$, which implies $\Pr(|\varepsilon_i| > M) < \kappa$. Moreover, since $\hat{\varepsilon}_{N,i} - \varepsilon_i = o_p(1)$, there exists an $N_0 > 0$, possibly depending on $i$, such that $\Pr(|\hat{\varepsilon}_{N,i} - \varepsilon_i| \geq M/2) < \kappa/2$ if $N > N_0$. Thus, under the latter condition

$$\Pr(|\hat{\varepsilon}_{N,i}| > M) = \Pr(|\varepsilon_i + (\hat{\varepsilon}_{N,i} - \varepsilon_i)| > M) \leq \Pr(|\varepsilon_i| + |\hat{\varepsilon}_{N,i} - \varepsilon_i| > M) = \Pr(|\varepsilon_i| + |\hat{\varepsilon}_{N,i} - \varepsilon_i| > M) \text{ and } |\hat{\varepsilon}_{N,i} - \varepsilon_i| < M/2 \right) + \Pr(|\varepsilon_i| + |\hat{\varepsilon}_{N,i} - \varepsilon_i| > M) \text{ and } |\hat{\varepsilon}_{N,i} - \varepsilon_i| \geq M/2 \right) \leq \Pr(|\varepsilon_i| > M/2) + \Pr(|\hat{\varepsilon}_{N,i} - \varepsilon_i| \geq M/2) < \kappa.$$ 

Now for $N > N_0$ and arbitrary $\xi > 0$,

$$\Pr(\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\hat{\varepsilon}_{N,i}) > \xi) = \Pr(\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\hat{\varepsilon}_{N,i}) > \xi \text{ and } |\hat{\varepsilon}_{N,i}| \leq M) + \Pr(\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\hat{\varepsilon}_{N,i}) > \xi \text{ and } |\hat{\varepsilon}_{N,i}| > M) \leq \Pr \left( \sup_{|\varepsilon| \leq M} |\hat{\eta}_N(\varepsilon) - \eta_0(\varepsilon)| > \xi \right) + \Pr(|\hat{\varepsilon}_{N,i}| > M) \right) \leq \Pr(\hat{\eta}_N(\varepsilon) - \eta_0(\varepsilon) > \xi) + \kappa,$n

hence by (15.17), $\limsup_{N \to \infty} \Pr(\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\hat{\varepsilon}_{N,i}) > \xi) \leq \kappa$, which by the arbitrariness of $\kappa$ and $\xi$ implies that

$$|\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\hat{\varepsilon}_{N,i})| = o_p(1)$$

for each $i$. Similarly, it follows that

$$|\hat{\eta}_N(\varepsilon_i) - \eta_0(\varepsilon_i)| = o_p(1)$$

for each $i$. Finally, it is not too hard to verify from the continuity of $\eta_0(\varepsilon)$ on $\mathbb{R}$ that

$$|\eta_0(\hat{\varepsilon}_{N,i}) - \eta_0(\varepsilon_i)| = o_p(1)$$

for each $i$. Combining these three results it follows that $\hat{\eta}_N(\hat{\varepsilon}_{N,i}) - \eta_0(\varepsilon_i) = o_p(1)$ for each $i \in \mathbb{N}$.

\footnote{As is well-known, convergence in probability is equivalent to a.s. convergence along a}
15.7. Miscellaneous derivations

Derivation of (6.16)

It follows from (6.14) that

\[ c_{k,m}(w) = \int_{0}^{w} d_{k,m}(x) \, dx \]

\[ = \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(-1)^{i+j}}{i!j!} \int_{0}^{w} x^{i+j} \exp(-x) \, dx \]

\[ = \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(-1)^{i+j}}{i!j!} (i+j)! \left( 1 - \exp(-w) \sum_{\ell=0}^{i+j} \frac{w^\ell}{\ell!} \right) \]

\[ = \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(i+j)!}{i!j!} (-1)^{i+j} \]

\[ \times \left( 1 - \exp(-w) \left( 1 + \sum_{\ell=1}^{i+j} \frac{w^\ell}{\ell!} \right) \right). \]

The second equality follows from the fact that for \( k \in \mathbb{N} \), by integration by parts,

\[ \int_{0}^{w} x^{k} \exp(-x) \, dx = \int_{0}^{w} x^{k} \frac{d(1-\exp(-x))}{dx} \, dx \]

\[ = w^{k}(1-\exp(-w)) - k \int_{0}^{w} x^{k-1} (1-\exp(-x)) \, dx \]

\[ = k \int_{0}^{w} x^{k-1} \exp(-x) \, dx - w^{k} \exp(-w), \]

so that by backwards substitution, with \( m \leq k \),

\[ \int_{0}^{w} x^{k} \exp(-x) \, dx = k(k-1) \int_{0}^{w} x^{k-2} \exp(-x) \, dx - (kw^{k-1} + w^{k}) \exp(-w) \]

further subsequence of an arbitrary subsequence. See for example Bierens (2004, Ch.6). Let \( N_{k} \) be such a further subsequence, so that \( \hat{U}_{N_{k},i} - U_{i} = o_{p}(1) \) is equivalent to \( \hat{U}_{N_{k},i} - U_{i} \xrightarrow{a.s.} 0 \) 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 \( \lim_{k \to \infty} \gamma_{0}(\hat{U}_{N_{k},i}(\omega)) = \gamma_{0}(U_{i}(\omega)) \) for all \( \omega \in \Omega \setminus N_{0} \), which implies that \( \gamma_{0}(\hat{U}_{N,i}) = \gamma_{0}(U_{i}) + o_{p}(1) \).

61
\[
\equiv \frac{k!}{(k-2)!} \int_0^w x^{k-2} \exp(-x) \, dx - k! \sum_{\ell=0}^{1} \frac{w^{k-\ell}}{(k-\ell)!} \exp(-w) \\
= \frac{k!}{(k-m)!} \int_0^w x^{k-m} \exp(-x) \, dx - k! \sum_{\ell=0}^{m-1} \frac{w^{k-\ell}}{(k-\ell)!} \exp(-w) \\
= k! \int_0^w \exp(-x) \, dx - k! \sum_{\ell=0}^{k-1} \frac{w^{k-\ell}}{(k-\ell)!} \exp(-w) \\
= k! \left( 1 - \exp(-w) \left( 1 + \sum_{\ell=1}^{k} \frac{w^\ell}{\ell!} \right) \right),
\]
where the result for \( m \) follows by induction.

**Derivation of (6.21)**

\[
b_{k,m}(t) = \int_0^\infty \exp(-t \cdot w) a_{k,m}(w) \, dw \\
= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(-1)^{i+j}}{i! j!} \int_0^\infty w^{i+j} \exp(-(1+t)w) \, dw \\
= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \frac{(-1)^{i+j}}{i! j!} \left( \frac{1}{1+t} \right)^{i+j+1} \int_0^\infty w^{i+j} \exp(-w) \, dw \\
= \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \left( \frac{i+j}{i} \right) (-1)^{i+j} \left( \frac{1}{1+t} \right)^{i+j+1},
\]
where the last equality follows straightforwardly from the easy equality \( \int_0^\infty w^\ell \exp(-w) \, dw = \ell! \), \( \ell \in \mathbb{N} \).

**Derivation of (7.2)**

Recall from (5.1) that
\[
f_0(w) = \frac{w^2 \exp(-w/\gamma_0)}{2\gamma_0^3}, \quad \gamma_0 = 1/(2\sqrt{3}), \quad w \geq 0.
\]

Then using the closed form expression (6.2) for \( \rho_m(w) \), and denoting \( \lambda_0 = (1 + \)
1/\gamma_0)/2 = 0.5 + \sqrt{3}, the integrals in (7.1) can be computed exactly as

\begin{align*}
\int_0^\infty \rho_m(w) \exp(-w/2)\sqrt{f_0(w)}dw &= \frac{1}{\sqrt{2}\gamma_0^{3/2}} \sum_{\ell=0}^m \left(\frac{m}{\ell}\right) \frac{(-1)^\ell}{\ell!} \int_0^\infty w^{\ell+1} \exp(-\lambda_0 w)dw \\
&= \frac{1}{\sqrt{2}\gamma_0^{3/2}} \sum_{\ell=0}^m \left(\frac{m}{\ell}\right) \frac{(-1)^\ell}{\ell!} \lambda_0^{2-\ell} \int_0^\infty w^{\ell+1} \exp(-w)dw \\
&= \frac{1}{\sqrt{2}\gamma_0^{3/2}} \sum_{\ell=0}^m \left(\frac{m}{\ell}\right) (-1)^\ell (\ell + 1) \lambda_0^{2-\ell}
\end{align*}

and similarly,

\begin{align*}
\int_0^\infty \rho_0(w) \exp(-w/2)\sqrt{f_0(w)}dw &= \int_0^\infty \exp(-w/2)\sqrt{f_0(w)}dw = \frac{1}{\sqrt{2}\gamma_0^{3/2}} \lambda_0^{-2}.
\end{align*}

**Computation of (9.15)**

Let

\[ f(w) = I(w > s)f_0(w - s), \quad s > 0, \]

where \( f_0(w) \) is the Gamma density (5.1). The integrals in (7.1) with \( f_0(w) \) replaced by \( f(w) \) and again \( \lambda_0 = (1 + 1/\gamma_0)/2 = 0.5 + \sqrt{3} \), now become

\begin{align*}
\int_0^\infty \rho_m(w) \exp(-w/2)\sqrt{f(w)}dw &= \int_0^\infty \rho_m(w) \exp(-w/2)I(w > s)\sqrt{f_0(w - s)}dw \\
&= \int_s^\infty \rho_m(w) \exp(-w/2) \frac{(w - s)\exp(-0.5(w - s)/\gamma_0)}{\sqrt{2}\gamma_0^{3/2}}dw \\
&= \int_0^\infty \rho_m(w + s) \exp(-(w + s)/2) \frac{w\exp(-0.5w/\gamma_0)}{\sqrt{2}\gamma_0^{3/2}}dw \\
&= \exp(-s/2) \frac{1}{\sqrt{2}\gamma_0^{3/2}} \int_0^\infty \rho_m(w + s)w \exp(-\lambda_0 w)dw
\end{align*}
\[ \begin{align*}
&= \exp(-s/2) \sqrt{2/\gamma_0^{3/2}} \sum_{\ell=0}^{m} \left( \frac{m}{\ell} \right) \frac{(-1)^{\ell}}{\ell!} \int_0^\infty (w+s)^{\ell} w \exp(-\lambda_0 w) dw \\
&= \exp(-s/2) \sqrt{2/\gamma_0^{3/2}} \sum_{\ell=0}^{m} \left( \frac{m}{\ell} \right) \frac{(-1)^{\ell}}{\ell!} \sum_{k=0}^{\ell} \left( \frac{\ell}{k} \right) s^{\ell-k} \int_0^\infty w^{k+1} \exp(-\lambda_0 w) dw \\
&= \exp(-s/2) \sqrt{2/\gamma_0^{3/2}} \sum_{\ell=0}^{m} \left( \frac{m}{\ell} \right) \frac{(-1)^{\ell}}{\ell!} \sum_{k=0}^{\ell} \left( \frac{\ell}{k} \right) \lambda_0^{-2-k} s^{\ell-k} \int_0^\infty w^{k+1} \exp(-w) dw \\
&= \exp(-s/2) \sqrt{2/\gamma_0^{3/2}} \sum_{\ell=0}^{m} \left( \frac{m}{\ell} \right) \frac{(-1)^{\ell}}{\ell!} \sum_{k=0}^{\ell} \left( \frac{\ell}{k} \right) s^{\ell-k} \lambda_0^{-k} (k+1)!
\end{align*} \]

and similarly,

\[ \begin{align*}
&\int_0^\infty \rho_0(w) \exp(-w/2) \sqrt{f(w)} dw \\
&= \int_0^\infty \exp(-w/2) I(w>s) \sqrt{f_0(w-s)} dw \\
&= \int_s^\infty \exp(-w/2) \int_s^\infty (w-s) \exp(-0.5(w-s)/\gamma_0) \sqrt{2/\gamma_0^{3/2}} dw \\
&= \frac{\exp(-s/2)}{\sqrt{2/\gamma_0^{3/2}}} \int_0^\infty \exp(-w/2) w \exp(-0.5w/\gamma_0) \sqrt{2/\gamma_0^{3/2}} dw \\
&= \frac{\exp(-s/2)}{\sqrt{2/\gamma_0^{3/2}}} \int_0^\infty w \exp(-\lambda_0 w) dw \\
&= \frac{\exp(-s/2)}{\lambda_0^2 \sqrt{2/\gamma_0^{3/2}}} \int_0^\infty w \exp(-w) dw \\
&= \frac{\exp(-s/2)}{\lambda_0^2 \sqrt{2/\gamma_0^{3/2}}}
\end{align*} \]

Thus,

\[ \delta_m(s) = \sum_{\ell=0}^{m} \left( \frac{m}{\ell} \right) \frac{(-1)^{\ell}}{\ell!} \sum_{k=0}^{\ell} \left( \frac{\ell}{k} \right) s^{\ell-k} \lambda_0^{-k} (k+1)! \\
= \sum_{\ell=0}^{m} (-1)^{\ell} \left( \frac{m}{\ell} \right) \sum_{k=0}^{\ell} \frac{(k+1)!}{k!(\ell-k)!} s^{\ell-k} \lambda_0^{-k} \]

64
\[ 
= \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{\ell} \sum_{k=0}^{\ell} \frac{k+1}{(\ell-k)!} \lambda_0^{-k} s^{\ell-k}, \quad m \in \mathbb{N}.
\]
16. Appendix B: Computational issues

16.1. Computation of the objective function

After some tedious but elementary calculations it can be shown that

\[ Q_N(\alpha, \sigma^2, \delta) = \int_0^1 \left( \Psi_N(t|\delta) - \alpha t - \sigma^2 t^2/2 \right)^2 dt \]

\[ = \int_0^1 (t \Psi_N(t|\delta))^2 dt - 3 \left( \int_0^1 t \Psi_N(t|\delta) dt \right)^2 \]

\[ - 80 \left( \int_0^1 t^2 \Psi_N(t|\delta) dt - \frac{3}{4} \int_0^1 t \Psi_N(t|\delta) dt \right)^2 \]

\[ + \frac{1}{3} \left( \alpha - \left( \int_0^1 t \Psi_N(t|\delta) dt - \frac{3}{8} \sigma^2 \right) \right)^2 \]

\[ + \frac{1}{320} \left( \sigma^2 - 160 \int_0^1 t^2 \Psi_N(t|\delta) dt + 120 \int_0^1 t \Psi_N(t|\delta) dt \right)^2 \]

Thus, \( Q_N(\alpha, \sigma^2, \delta) \) involves three types of integrals, namely

\[ \int_0^1 \Psi_N(t|\delta)^2 dt, \int_0^1 t \Psi_N(t|\delta) dt \text{ and } \int_0^1 t^2 \Psi_N(t|\delta) dt, \]

where \( \Psi_N(t|\delta) = \Upsilon(t) - \ln(\mathcal{L}(t|\delta)) \). However, due to the logs in \( \Upsilon_N(t) \) and \( \ln(\mathcal{L}(t|\delta)) \) these integrals do not have closed form expressions. On the other hand, it follows from Theorem 2 that, at least asymptotically, \( \Upsilon_N(t) \) is a smooth function, being in the limit the sum of a quadratic function and the strictly monotonic decreasing function \( \ln(E[\exp(-tW)]) \). Similarly, \( \ln(\mathcal{L}(t|\delta)) \) is a strictly monotonic decreasing function in \( t \), and therefore is smooth as well. The same applies to \( \Psi_N(t|\delta) = \Upsilon_N(t) - \ln(\mathcal{L}(t|\delta)) \). Therefore, it seems that the integrals involved can be approximated very close by the Legendre-Gauss quadrature, to be reviewed briefly in the next subsection.

16.2. The Legendre-Gauss quadrature

One of the standard numerical integration procedures is the Legendre-Gauss quadrature\(^{30} \) [LGQ hereafter] to approximate integrals of the type \( \int_{-1}^1 \phi(x) dx \),

\(^{30}\)See for example Press et al (1989), Abbott (2005) and the link
for square integrable real functions \(\phi(x)\) on \([-1, 1]\), by
\[
\int_{-1}^{1} \phi(x)\,dx \approx \sum_{\ell=1}^{L} w_\ell \phi(x_\ell)
\] (16.1)
for some \(L \geq 2\), where the abscissas \(x_\ell\) and the weights \(w_\ell > 0\) depend on \(L\) but not on \(\phi\). Therefore, for each \(L\) the pairs \((w_\ell, x_\ell)\) can be tabulated. Moreover, the LGQ approximation (16.1) is exact if \(\phi(x)\) itself is a polynomial of order \(2L-1\) or less. The latter implies that
\[
\sum_{\ell=1}^{L} w_\ell = \int_{-1}^{1} 1\,dx = 2.
\]

Since \(\int_{0}^{1} \phi(u)\,du = 0.5 \int_{-1}^{1} \phi((x + 1)/2)\,dx\), the LGQ approximation (16.1) also applies to square integrable real functions on the unit interval, as
\[
\int_{0}^{1} \phi(u)\,du \approx 0.5 \sum_{\ell=1}^{L} w_\ell \phi((x_\ell + 1)/2) = \sum_{\ell=1}^{L} b_\ell \phi(a_\ell),
\] (16.2)
say, for some fixed but not too small an \(L\), where
\[
a_\ell = (x_\ell + 1)/2, \quad b_\ell = w_\ell/2, \quad \ell = 1, 2, \ldots, L.
\] (16.3)

As an example of how accurate the LGQ approximation can be, consider the Laplace transform of the \(\chi^2_2\) distribution, i.e., \(L(t|\chi^2_2) = 1/(1+2t)\). It is not hard to verify that
\[
\int_{0}^{1} \ln \left( L(t|\chi^2_2) \right) \,dt = 1 - 1.5 \ln(3) = -0.64791843302165.
\]

Next, let us compute this integral via the LGQ approach, as \(\sum_{\ell} b_\ell \ln(1/(1+2a_\ell))\), for \(L = 5, 10, 20, 30, 40, 50\). The results are presented in Table B.1.

<table>
<thead>
<tr>
<th>(L)</th>
<th>Accuracy of the LGQ approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(-0.64791843302165)</td>
</tr>
<tr>
<td>10</td>
<td>(-0.532160589540430)</td>
</tr>
<tr>
<td>20</td>
<td>(-0.647918433002425)</td>
</tr>
<tr>
<td>30</td>
<td>(-0.64791843302164)</td>
</tr>
<tr>
<td>40</td>
<td>(-0.647918433002165)</td>
</tr>
<tr>
<td>50</td>
<td>(-0.647918433002165)</td>
</tr>
</tbody>
</table>

- \(http://mathworld.wolfram.com/Legendre-GaussQuadrature.html\)
- \(^{31}\)These tables are given in the link https://pomax.github.io/bezierinfo/legendre-gauss.html, for \(L = 2, 3, \ldots, 64\).
Moreover, recall from (6.20) that we need to compute the matrices where the
are due to the monotonicity of the integrand
percentage error is about \(-18\%\). Our guess is that the favorable results for \(L \geq 10\) are due to the monotonicity of the integrand \(\ln(1/(1 + 2t))\).

### 16.3. Computational efficiency

We will now address the question how to compute the LGQ approximations

\[
\int_0^1 \Psi_N(t|\pi_n\delta)^2 \, dt \approx \sum_{\ell=1}^L b_\ell \Psi_N(a_\ell|\pi_n\delta)^2 = \sum_{\ell=1}^L b_\ell (\Upsilon_N(a_\ell) - \ln(L(a_\ell|\pi_n\delta)))^2,
\]

\[
\int_0^1 t\Psi_N(t|\pi_n\delta) \, dt \approx \sum_{\ell=1}^L b_\ell a_\ell \Psi_N(a_\ell|\pi_n\delta) = \sum_{\ell=1}^L b_\ell a_\ell (\Upsilon_N(a_\ell) - \ln(L(a_\ell|\pi_n\delta)))
\]

\[
\int_0^1 t^2\Psi_N(t|\pi_n\delta) \, dt \approx \sum_{\ell=1}^L b_\ell a_\ell^2 \Psi_N(a_\ell|\pi_n\delta) = \sum_{\ell=1}^L b_\ell a_\ell^2 (\Upsilon_N(a_\ell) - \ln(L(a_\ell|\pi_n\delta)))
\]

with minimal computation time.

Clearly, we need to compute the sequence \(\Upsilon_N(a_\ell), \ell = 1, 2, ..., L\), only once. Moreover, recall from (6.20) that

\[
L(t|\pi_n\delta) = \frac{\xi_{n+1}(\delta)'B_{n+1}(t)\xi_{n+1}(\delta)}{\xi_{n+1}(\delta)'\xi_{n+1}(\delta)}, \quad \xi_{n+1}(\delta)' = (1, \delta_1, \delta_2, \ldots, \delta_n),
\]

where \(B_{n+1}(t)\) is a \((n + 1) \times (n + 1)\) matrix-valued function with \((k, m)\) elements defined in (6.21). Thus, to compute the \(L(a_\ell|\pi_n\delta))\)'s for different \(\delta\)'s efficiently we need to compute the matrices \(B_{n+1}(a_\ell)\) for \(\ell = 1, 2, \ldots, L\) only once.\(^{32}\)

Similarly,

\[
L'(t|\pi_n\delta) = \frac{\xi_{n+1}(\delta)'B_{n+1}'(t)\xi_{n+1}(\delta)}{\xi_{n+1}(\delta)'\xi_{n+1}(\delta)}
\]

\[
L''(t|\pi_n\delta) = \frac{\xi_{n+1}(\delta)'B_{n+1}''(t)\xi_{n+1}(\delta)}{\xi_{n+1}(\delta)'\xi_{n+1}(\delta)}
\]

where the \((k, m)\) elements of \(B_{n+1}'(t)\) and \(B_{n+1}''(t)\) are

\[
b'_{k,m}(t) = \sum_{i=0}^k \sum_{j=0}^m \binom{k}{i} \binom{m}{j} (-1)^{i+j+1} (i+j+1) \left( \frac{1}{1+t} \right)^{i+j+2},
\]

\(^{32}\)And then store them in a three dimensional \((n + 1) \times (n + 1) \times L\) common array, which is possible in Visual Basic 5 and possibly in other programming languages as well.

68
\[ b''_{k,m}(t) = \sum_{i=0}^{k} \sum_{j=0}^{m} \binom{k}{i} \binom{m}{j} \binom{i + j}{i} (-1)^{i+j+2} (i + j + 1)(i + j + 2) \left( \frac{1}{1 + t} \right)^{i+j+3}, \]

respectively. Thus, to compute

\[
\overline{\psi}_N(\delta) = \min_{i=1,2,\ldots,K} (\psi'_N(i/K|\delta) - (i/K) \cdot \psi''_N(i/K|\delta)) \\
\overline{\sigma}_N^2(\delta) = \max_{i=1,2,\ldots,K} \psi''_N(i/K|\delta)
\]

efficiently for different \( \delta \)'s it suffices to compute \( \psi'_N(i/K), \psi''_N(i/K) \) and the matrices \( B'_{n+1}(i/K) \) and \( B''_{n+1}(i/K) \) for \( i = 1, 2, \ldots, K \) only once.