Supplementary Material on "Semiparametric Estimation of Random Coefficients in Structural Economic Models"

Stefan Hoderlein* Lars Nesheim † Anna Simoni‡
Boston College University College London CNRS - CREST

September 20, 2016

This document presents several appendices that contain supplementary results to be published online alongside the main paper, "Semiparametric Estimation of Random Coefficients in Structural Economic Models". All references to equation numbers and to section numbers are references to equations and sections in the main paper. The main paper contains Appendices A and B. This document begins with Appendix C.

C Compactness

The following proposition shows that under Assumptions 1-6 the operator T defined in (4.2) is compact with infinite dimensional range. As discussed in Section 4 in the paper, compactness of the operator is useful because then T admits a SVD.

Proposition 4. *Let T be the operator defined in (4.2) with domain $L^2_{\pi_\theta}$ and let Assumptions 1 - 6 be satisfied. If $f_{C|WZ\theta}/\pi_\theta$ is square integrable with respect to $\pi_\theta \times \pi_{cz}$ then $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$ and $T : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ is an a.s. bounded and compact operator.*

The proof is detailed in Appendix F.2 below.

*Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com.
†Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK, email: l.nesheim@ucl.ac.uk
‡CREST, 15 Boulevard Gabriel Péri, 92240 Malakoff (France), email: simoni.anna@gmail.com

D Identification and completeness

In addition to the large class of functions that satisfy the sufficient conditions for identification given in Proposition 3, we provide here further examples of families $\mathcal{F}_{C|WZ\theta}$ for which the corresponding $\mathcal{F}_{\theta|CWZ}$ is $\mathcal{T}$-complete.

Additively-closed one-parameter family of distributions. Let $\Theta = \mathbb{R}_+$ and $\mathcal{F}_{C|WZ\theta}$ be *additively closed*. That is, $\forall f_{C|WZ\theta}, h_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$ and $\forall \theta_1, \theta_2 \in \Theta$,

$$f_{C|WZ\theta}(c, w, z, \theta_1) * h_{C|WZ\theta}(c, w, z, \theta_2) = f_{C|WZ\theta}(c, w, z, \theta_1 + \theta_2),$$

where $*$ denotes the convolution operation. Then, $\mathcal{F}_{\theta|CWZ}$ is $\mathcal{T}-$complete. Some distributions that belong to the additively-closed one-parameter family, and that are relevant for our application, are the following, see Teicher (1961).

- Gamma distribution: $f_{C|WZ\theta} = \frac{g(z,w)^\theta}{\Gamma(\theta)} c^{\theta-1} e^{-g(z,w)c}$, $c > 0$, $g(z, w) > 0$, $\theta > 0$ or $f_{C|WZ\theta} = \frac{\theta^{g(z,w)}}{\Gamma(g(z,w))} c^{g(z,w)-1} e^{-\theta c}$, $c > 0$, $g(z, w) > 0$, $\theta > 0$.
- Uniform distribution with support depending on θ: $f_{C|WZ\theta} = \mathcal{U}[\theta - g(Z, W), \theta + g(Z, W)]$, where $g(\cdot, \cdot)$ is some positive and bounded function of (Z, W). Therefore,

  $$f_{C|WZ\theta} = \frac{1}{2g(Z, W)} 1\{\theta - g(Z, W) < c < \theta + g(Z, W)\}.$$

  However, if $f_{C|WZ\theta}$ has a uniform distribution with support that does not depend on θ then, $f_{\theta|W}$ is not identified.

Location-scale one-parameter family of distributions. Let $\Theta = \mathbb{R}_+$ and $\mathcal{F}_{C|WZ\theta}$ be the one-parameter family induced by $f_{C|WZ}$ via location or scale changes. That is, $\forall f_{C|WZ\theta} \in \mathcal{F}_{C|WZ\theta}$, $f_{C|WZ\theta}(c, w, z, \theta) = f_{C|WZ}(c - \theta, w, z)$ or $f_{C|WZ\theta}(c, w, z, \theta) = f_{C|WZ}(c\theta, w, z)$. For the location (resp. scale) family, if the conditional characteristic function of C (resp. $\log C$), given (W, Z), does not vanish a.s. in some non-degenerate real interval, then the $f_{\theta|W}$ is identified, see Teicher (1961).

D.1 Identification without nuisance unobservables

In this section we briefly describe the case where we do not have ε so that $f_{C|WZ\theta}$ cannot be recovered as in Theorem 1. This is relevant in models where all the unobservable variables are of interest so ε is included in θ. In our setup, this implies that the general structural model (3.1) reduces to

$$\Psi(C, W, Z, \theta) = 0 \qquad a.s. \tag{D.1}$$

and Assumption 1 is replaced by the following one.

Assumption 1′. *The random element (C, W, Z, θ) satisfies a structural economic model*

$$\Psi(C, W, Z, \theta) = 0 \quad a.s. \tag{D.2}$$

*where Ψ is a **known** Borel measurable real-valued function. We assume that (D.2) has a unique global solution in terms of C:*

$$C = \varphi(W, Z, \theta), \quad a.s. \tag{D.3}$$

where $\varphi : \mathbb{R}^{k+l+d} \to \mathbb{R}$ is a Borel-measurable function.

Indeed, even in this setup where φ is not strictly monotonic in θ and θ is multivariate, we can characterize the structural *pdf* $f_{\theta|W}$ as a solution to a constrained functional equation. Let $F_{C|WZ}$ be the cumulative distribution function associated with $P_{C|WZ}$ and assumed to be in $L^2_{\pi_{cz}}$ for every $w \in \mathcal{W}$. Then, we have the following analog to Theorem 1.

Theorem 8. *Let Assumptions 1′ and 5 be satisfied. If $P_{\theta|W}$ admits a pdf $f_{\theta|W}$ with respect to the Lebesgue measure, then $f_{\theta|W}$ is a solution of:*

$$F_{C|WZ}(c, w, z) = Sf_{\theta|W}(\theta, w) \quad \textit{subject to} \quad f_{\theta|W} \in \mathcal{F}_{\theta|W}, \qquad a.s. \tag{D.4}$$

where S is a linear operator defined as

$$Sh = \int_\Theta 1_{\{\varphi(w,z,\theta)\leq c\}}(\theta)h(\theta, w)d\theta, \quad \forall h \in L^2_{\pi_\theta}. \tag{D.5}$$

Proof. Equations (D.4)-(D.5) follow from the fact that, under Assumption 1′, $F_{C|WZ}(c, w, z) = \mathbb{E}\left[1_{\{\varphi(w,z,\theta)\leq c\}}(\theta)\middle| W, Z\right] = \int 1_{\{\varphi(w,z,\theta)\leq c\}}(\theta)dP_{\theta|W,Z}(\theta, w, z)$ and from Assumption 5. ■

The kernel of the operator S is $\frac{1\{\varphi(w,z,\theta)\leq c\}}{\pi_\theta(\theta)}$ and the adjoint S^* is given in the following proposition:

Proposition 5 (Adjoint of S)**.** *Let S be the operator defined in (D.5). Assume that $S : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ is bounded. Then, the operator S^* defined as: $\forall \psi \in L^2_{\pi_{cz}}$,*

$$S^*h = \int_\mathcal{C}\int_\mathcal{Z} 1_{\{\varphi(w,z,\theta)\leq c\}}(\theta)\frac{\pi_{cz}(c, z)}{\pi_\theta(\theta)}h(c, w, z)dcdz.$$

exists and is the adjoint of S. The operator $S^* : L^2_{\pi_{cz}} \to L^2_{\pi_\theta}$ is bounded and linear.

The proof is similar to the proof of Proposition 1 and is omitted. Note that when there are nuisance unobservables ε, the estimating equation (4.3) can be trivially recovered from (D.4) by differentiating with respect to c. If $\int_{\mathcal{C}\times\mathcal{Z}}\int_\Theta \frac{1}{\pi_\theta}d\theta\pi_{cz}dcdz < \infty$, then the bounded operator $S : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ is compact.

Identification of $f_{\theta|W}$ depends on injectivity of $S|_{\mathfrak{D}}$ which, in turn, depends on the exogenous variation in Z. The estimation procedure for this case is the same as that one proposed in Section 5 with the operator T replaced by S. The rate of the mean integrated squared error will improve since $F_{C|WZ}$ can be estimated at a better rate than $f_{C|WZ}$. Moreover, the degree of ill-posedness will not be as severe as in the case where the kernel of T is exponential.

E Case with non-random parameters: Iterative two-step method

In this section we describe the two-step estimator in the case in which some components of θ are deterministic as described in Section 5.3. This is an iterative algorithm similar to that proposed in Heckman & Singer (1984). The algorithm is as follows:

I. For a given $\theta_1^{(j)}$ compute the *indirect Tikhonov regularized estimator* of $f_{\theta_2|W}$ using the two-step procedure described in Section 5.1. That is, in the *first step* solve the minimization problem

$$\hat{f}^{\alpha}_{\theta_2|W(j)} = \arg\min_{h \in L^2_{\pi_{\theta_2}}} \left\{ ||T_{\theta_1^{(j)}} h - \hat{f}_{C|WZ}||^2 + \alpha ||h||^2 \right\}$$

and in the second step compute the metric projection of $\hat{f}^{\alpha}_{\theta_2|W(j)}$ onto the set $\mathcal{F}_{\theta|W}$ as

$$\mathcal{P}_c \hat{f}^{\alpha}_{\theta_2|W(j)} = \max\left\{0, \hat{f}^{\alpha}_{\theta_2|W(j)} - \frac{c}{\pi_{\theta_2}}\right\} \tag{E.1}$$

where c is such that $\int_\Theta \mathcal{P}_c \hat{f}^{\alpha}_{\theta_2|W(j)} d\theta = 1$. Fix $\hat{f}^{(j)}_{\theta_2|W} = \mathcal{P}_c \hat{f}^{\alpha}_{\theta_2|W(j)}$.

II. For a given $\hat{f}^{(j)}_{\theta_2|W}$ compute $\theta_1^{(j+1)}$ by solving the nonlinear least-squares problem:

$$\theta_1^{(j+1)} = \arg\min_{\theta_1 \in \Theta_1} \left(||T_{\theta_1} \hat{f}^{(j)}_{\theta_2|W}(\theta_2, w) - \hat{f}_{C|WZ}||^2 + \alpha ||\hat{f}^{(j)}_{\theta_2|W}||^2 \right).$$

Then, iterate steps I and II until convergence. The algorithm should be run using different starting values for θ_1 to avoid convergence to a local optimum.

F Proofs of minor results

F.1 Proof of Proposition 1

By definition, the adjoint operator T^* of the bounded linear operator T satisfies: $\forall h \in L^2_{\pi_\theta}$, $\forall \psi \in L^2_{\pi_{cz}}$, $\langle Th, \psi \rangle = \langle h, T^*\psi \rangle$. Thus,

$$\begin{aligned}
\langle Th, \psi \rangle &= \int_{\mathcal{C}} \int_{\mathcal{Z}} (Th)(c, w, z)\psi(c, z)\pi_{cz}(c, z)dcdz \\
&= \int_{\mathcal{C}} \int_{\mathcal{Z}} \int_{\Theta} f_{C|WZ\theta}(c, w, z, \theta)h(\theta)d\theta\psi(c, z)\pi_{cz}(c, z)dcdz \\
&= \int_{\Theta} h(\theta)\pi_\theta(\theta) \int_{\mathcal{C}} \int_{\mathcal{Z}} f_{C|WZ\theta}(c, w, z, \theta)\psi(c, z)\frac{\pi_{cz}(c, z)}{\pi_\theta(\theta)}dcdzd\theta = \langle h, T^*\psi \rangle
\end{aligned}$$

where the third equality follows from the Fubini's theorem. Existence and linearity follow from the Riesz representation theorem. Boundedness of T^* follows from the boundedness of T since $||T^*|| = ||T||$.

F.2 Proof of Proposition 4

We first show that $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$. By the Cauchy-Schwarz inequality, $\forall w \in \mathcal{W}$ and $\forall h \in L^2_{\pi_{cz}}$:

$$\begin{aligned}
||Th||^2 &= \int_{\mathcal{C}} \int_{\mathcal{Z}} \left\langle \frac{f_{C|WZ\theta}}{\pi_\theta}, h \right\rangle^2 \pi_{cz}(c, z)dcdz \\
&\leq \int_{\mathcal{C}} \int_{\mathcal{Z}} \left|\left| \frac{f_{C|WZ\theta}}{\pi_\theta} \right|\right|^2 ||h||^2 \pi_{cz}(c, z)dcdz \\
&= ||h||^2 \int_{\mathcal{C}} \int_{\mathcal{Z}} \int_{\Theta} \frac{f^2_{C|WZ\theta}}{\pi_\theta} \pi_{cz} d\theta dcdz.
\end{aligned} \tag{F.1}$$

The expression is finite if the multiple integral is bounded. This is shown below in the second part of the proof. Thus, after showing this we establish that $\mathcal{R}(T) \subset L^2_{\pi_{cz}}$.

Next, we show compactness of T. This can be shown by showing that T is Hilbert-Schmidt. An integral operator from $L^2_{\pi_\theta}$ to $L^2_{\pi_{cz}}$ is Hilbert-Schmidt if its kernel is square integrable with respect to $\pi_\theta \times \pi_{cz}$. An Hilbert-Schmidt operator is bounded and compact. Under the conditions of the proposition we compute $\int_{\mathcal{C}} \int_{\mathcal{Z}} \int_{\Theta} \frac{f^2_{C|WZ\theta}}{\pi^2_\theta} \pi_\theta \pi_{cz}$ and show that it is bounded:

$$\begin{aligned}
&\int_{\mathcal{C}} \int_{\mathcal{Z}} \int_{\Theta} \frac{f^2_{C|WZ\theta}}{\pi^2_\theta} \pi_\theta \pi_{cz} \\
= &\int_{\mathcal{C}} \int_{\mathcal{Z}} \int_{\Theta} \left[\sum_{i=1}^{s} f_{\varepsilon|WZ\theta}(\varphi_i^{-1}(w, z, \theta, c), w, z, \theta) \left| \partial_c \varphi_i^{-1}(w, z, \theta, c) \right| 1_{\mathcal{C}_i}(c) \right]^2 \frac{\pi_\theta(\theta)\pi_{cz}(c, z)}{\pi^2_\theta} d\theta dcdz \\
\leq &\int_{\Theta} \int_{\mathcal{Z}} 2^{s-1} \sum_{i=1}^{s} \int_{\mathcal{C}} f^2_{\varepsilon|WZ\theta}(\varphi_i^{-1}(w, z, \theta, c), w, z, \theta) \left| \partial_c \varphi_i^{-1}(w, z, \theta, c) \right|^2 1_{\mathcal{C}_i}(c) \frac{\pi_{cz}(c, z)}{\pi_\theta} dcdzd\theta \\
= &\int_{\Theta} \int_{\mathcal{Z}} 2^{s-1} \sum_{i=1}^{s} \int_{\mathcal{E}_i} f^2_{\varepsilon|WZ\theta}(\varepsilon_i, w, z, \theta) \left| \partial_{\varepsilon_i} \varphi(w, z, \theta, \varepsilon_i) \right|^{-1} \frac{\pi_{cz}(\varphi(w, z, \theta, \varepsilon_i), z)}{\pi_\theta(\theta)} d\varepsilon_i dzd\theta \\
< &\ \infty
\end{aligned}$$

where the first inequality follows from the Fubini's theorem and the Cauchy-Schwarz's inequality and the second equality follows from the change of variable $\varphi_i^{-1}(w,z,\theta,c)=\varepsilon_i$. The final inequality follows from Assumption 6. This result shows that $\mathcal{R}(T)\subset L^2_{\pi_{cz}}$, and that T is Hilbert-Schmidt and then bounded and compact.

G Technical lemmas

Lemma 3. *Let the assumptions of Corollary 1 be satisfied and $\hat{f}_{C|WZ}$ be as defined in* (5.7)*. Then,*

(i) $\left[\mathbb{E}(T^*\hat{f}_{C|WZ}-T^*f_{C|WZ})\right]^2=\mathcal{O}\left(\max\{h_n^4,h_d^4\}\right)$

(ii) $Var(T^*\hat{f}_{C|WZ})=\mathcal{O}\left[n^{-1}\left(\min\{h_n,h_d\}\right)^{-k}\right]$.

Proof. Note that $\hat{f}_{C|WZ}-f_{C|WZ}=\frac{1}{f_{WZ}}\left(\hat{f}_{CWZ}-f_{C|WZ}\hat{f}_{WZ}\right)\left[1-\left(\hat{f}_{WZ}-f_{WZ}\right)/\hat{f}_{WZ}\right]$. And, since $\left(\hat{f}_{WZ}-f_{WZ}\right)/\hat{f}_{WZ}=o_p(1)$ we can use the approximation $\hat{f}_{C|WZ}-f_{C|WZ}\simeq\frac{1}{f_{WZ}}\left(\hat{f}_{CWZ}-f_{C|WZ}\hat{f}_{WZ}\right)$. We start by showing result *(i)*.

Let t be a k-dimensional vector and v a l-dimensional vector. We use the notation $\overrightarrow{vt}=(v',t')$ and $\overrightarrow{uvt}=(u,v',t')$. Moreover, we let $p=k+l$ and let $D^2(h)$ be the Hessian matrix of a function h. We use a single integral symbol to denote the multiple integral either with respect to $dvdt$ or $dudvdt$. We start by computing the bias term $b(w,\theta)=\mathbb{E}\left(T^*\hat{f}_{C|WZ}-T^*f_{C|WZ}\right)$.

By standard Taylor series approximations we get: $b\left(w,\theta\right)\simeq T^*\frac{1}{f_{WZ}}\left[\mathbb{E}\left(\hat{f}_{CWZ}\right)-f_{C|WZ}\mathbb{E}\left(\hat{f}_{WZ}\right)\right]$ and then

$$
\begin{aligned}
b(w,\theta) &\simeq T^*\frac{1}{f_{WZ}}\left\{\left[\mathbb{E}\left(\hat{f}_{CWZ}\right)-f_{CWZ}\right]+f_{C|WZ}\left[f_{WZ}-\mathbb{E}\left(\hat{f}_{WZ}\right)\right]\right\};\\
\mathbb{E}\left(\hat{f}_{CWZ}\right)-f_{CWZ} &= \frac{h_n^2}{2}tr\left(D^2(f_{CWZ})\int\overrightarrow{uvt}'\overrightarrow{uvt}K(u,c)K(v,z)K(t,w)dudvdt\right)+o(h_n^2);\\
\mathbb{E}\left(\hat{f}_{WZ}\right)-f_{WZ} &= \frac{h_d^2}{2}tr\left(D^2(f_{WZ})\int\overrightarrow{vt}'\overrightarrow{vt}K(v,z)K(t,w)dvdt\right)+o(h_d^2);\\
b\left(w,\theta\right) &\simeq \int_{\mathcal{C}}\int_{\mathcal{Z}}\frac{f_{C|WZ\theta}}{f_{WZ}}\left[h_n^2tr\left(D^2(f_{CWZ})(c,w,z)\int\overrightarrow{uvt}'\overrightarrow{uvt}K(u,c)K(v,z)K(t,w)dudvdt\right)dcdz\right.\\
&\quad\left.-h_d^2tr\left(D^2(f_{WZ})(w,z)\int\overrightarrow{vt}'\overrightarrow{vt}K(v,z)K(t,w)dvdt\right)\right]\frac{\pi_{cz}(c,z)}{\pi_\theta(\theta)}dcdz+o\left(\max\{h_n^2,h_d^2\}\right)\\
&= h_n^2b_1(w,\theta)-h_d^2b_2(w,\theta)+o\left(\max\{h_n^2,h_d^2\}\right).
\end{aligned}
$$

Therefore, $b^2(w,\theta)=\mathcal{O}\left(\max\{h_n^4,h_d^4\}\right)$ which proves *(i)*.

Now consider the variance term (part *(ii)* of the Lemma).

$$
\begin{aligned}
Var\left(T^*\hat{f}_{C|WZ}\right) &= Var\left[T^*\left(\hat{f}_{C|WZ}-f_{C|WZ}\right)\right] \\
&\simeq Var\left[T^*\frac{1}{f_{WZ}}\left(\hat{f}_{CWZ}-f_{C|WZ}\hat{f}_{WZ}\right)\right] \\
&= Var\left(T^*\frac{\hat{f}_{CWZ}}{f_{WZ}}\right)+Var\left(T^*\frac{f_{C|WZ}\hat{f}_{WZ}}{f_{WZ}}\right) \\
&\quad -2Cov\left(T^*\frac{\hat{f}_{CWZ}}{f_{WZ}},T^*\frac{f_{C|WZ}\hat{f}_{WZ}}{f_{WZ}}\right).
\end{aligned}
$$

In the following we use the notation: $K_{h,i}(z,w)=K_h(z_i-z,z)K_h(w_i-w,w)$. We start by analysing the first term:

$$
\begin{aligned}
Var\left(T^*\frac{\hat{f}_{CWZ}}{f_{WZ}}\right) &= Var\left[\int_{\mathcal{Z}}\int_{\mathcal{C}}\frac{f_{C|WZ\theta}(c,w,z,\theta)}{f_{WZ}(w,z)nh_n^p}\sum_{i=1}^{n}\frac{K_h(c_i-c,c)}{h_n}K_{h,i}(z,w)\frac{\pi_{cz}(c,z)}{\pi_\theta}dcdz\right] \quad \text{(G.1)}\\
&= Var\left[\frac{1}{nh_n^k}\sum_{i=1}^{n}f_{C|WZ\theta}(c_i,w,z_i,\theta)\frac{\pi_{cz}(c_i,z_i)}{f_{WZ}(w,z_i)}\frac{K_h(w_i-w,w)}{\pi_\theta}\right]+o\left((nh_n^k)^{-1}\right) \\
&= \frac{1}{nh_n^{2k}}\int f^2_{C|WZ\theta}(c_i,w,z_i,\theta)\frac{\pi^2_{cz}(c_i,z_i)}{f^2_{WZ}(w,z_i)}\frac{K_h^2(w_i-w,w)}{\pi_\theta^2}f_{CWZ}(c_i,w_i,z_i)dc_idw_idz_i \\
&\quad -\frac{1}{nh_n^{2k}}\left[\int f_{C|WZ\theta}(c_i,w,z_i,\theta)\frac{\pi_{cz}(c_i,z_i)}{f_{WZ}(w,z_i)}\frac{K_h(w_i-w,w)}{\pi_\theta}f_{CWZ}(c_i,w_i,z_i)dc_idw_idz_i\right]^2 \\
&\quad +o\left(\frac{1}{nh_n^k}\right) \\
&= \frac{1}{nh_n^k}\int f^2_{C|WZ\theta}(c_i,w,z_i,\theta)\frac{\pi^2_{cz}(c_i,z_i)}{f_{WZ}(w,z_i)}\frac{\int K^2(t,w)dt}{\pi_\theta^2}f_{C|WZ}(c_i,w,z_i)dc_idz_i \\
&\quad +o\left((nh_n^k)^{-1}\right).
\end{aligned}
$$

Next,

$$
\begin{aligned}
Var\left(T^*\frac{f_{C|WZ}\hat{f}_{WZ}}{f_{WZ}}\right) &= Var\left(\int_{\mathcal{Z}}\int_{\mathcal{C}}\frac{f_{C|WZ\theta}(c,w,z,\theta)}{f_{WZ}(w,z)nh_d^p}\sum_{i=1}^{n}f_{C|WZ}(c,w,z)K_{h,i}(z,w)\frac{\pi_{cz}(c,z)}{\pi_\theta}dcdz\right) \quad \text{(G.2)}\\
&= Var\left(\frac{1}{nh_d^k}\sum_{i=1}^{n}\int_{\mathcal{C}}\frac{f_{C|WZ\theta}(c,w,z_i,\theta)}{f_{WZ}(w,z_i)}f_{C|WZ}(c,w,z_i)K_h(w_i-w,w)\frac{\pi_{cz}(c,z_i)}{\pi_\theta}dc\right) \\
&\quad +o\left(\frac{1}{nh_d^k}\right) \\
&= \frac{1}{nh_d^k}\int_{\mathcal{Z}}\left(\int_{\mathcal{C}}\frac{f_{C|WZ\theta}(c,w,z_i,\theta)}{f_{WZ}(w,z_i)}f_{C|WZ}(c,w,z_i)\pi_{cz}(c,z_i)dc\right)^2\times \\
&\qquad \frac{\int K^2(t,w)dt}{\pi_\theta^2}f_{WZ}(w,z_i)dz_i+o\left((nh_d^k)^{-1}\right),
\end{aligned}
$$

where the results are obtained by standard Taylor series approximations.

Finally, we have to compute the covariance term:

$$
\begin{aligned}
& Cov\left(T^*\frac{\hat{f}_{CWZ}}{f_{WZ}}, T^*\frac{f_{C|WZ}\hat{f}_{WZ}}{f_{WZ}}\right) \qquad \text{(G.3)}\\
= \; & \frac{1}{n^2 h_n^k h_d^k}\sum_{i=1}^{n} Cov\left(\begin{array}{c} \frac{f_{C|WZ\theta}(c_i,w,z_i,\theta)}{f_{WZ}(w,z_i)}K_h(w_i-w,w)\frac{\pi_{cz}(c_i,z_i)}{\pi_\theta}, \\ \int_{\mathcal{C}}\frac{f_{C|WZ\theta}(c,w,z_i,\theta)}{f_{WZ}(w,z_i)}K_h(w_i-w,w)\frac{\pi_{cz}(c,z_i)}{\pi_\theta}f_{C|WZ}(c,w,z_i)dc \end{array}\right)\\
& + o\left\{\left[n\left(\min\{h_n,h_d\}\right)^k\right]^{-1}\right\}\\
= \; & \frac{1}{nh_d^k}\int\left[\int_{\mathcal{C}} f_{C|WZ\theta}(c,w,z_i,\theta)f_{C|WZ}(c,w,z_i)\frac{\pi_{cz}(c,z_i)}{\pi_\theta}dc\right]^2\frac{1}{f_{WZ}(w,z_i)}dz_i\\
& \quad \times\int K(t,w)K\left(\frac{th_n}{h_d},w\right)dt + o\left((n(\min\{h_n,h_d\})^k)^{-1}\right).
\end{aligned}
$$

By putting (G.1), (G.2) and (G.3) together we obtain

$$
\begin{aligned}
& Var(T^*\hat{f}_{C|WZ}) \qquad \text{(G.4)}\\
\simeq \; & \int_{\mathcal{Z}}\frac{1}{nh_n^k}\left[\mathbb{E}\left(f_{C|WZ\theta}^2\pi_{cz}^2\,|w,z_i\right) + \frac{1}{nh_d^k}\mathbb{E}\left(f_{C|WZ\theta}\pi_{cz}\,|w,z_i\right)^2\right]\frac{\int K^2(t,w)dt}{f_{WZ}(w,z_i)\pi_\theta^2}dz_i\\
& -\frac{2}{nh_d^k}\int\frac{\mathbb{E}\left(f_{C|WZ\theta}\pi_{cz}\,|w,z_i\right)^2}{f_{WZ}(w,z_i)}dz_i\frac{\int K(t,w)K\left(\frac{th_n}{h_d}\right)dt}{\pi_\theta^2} + o\left(\frac{1}{n(\min\{h_n,h_d\})^k}\right)\\
= \; & \frac{1}{nh_n^k}V_1(w,\theta) + \frac{1}{nh_d^k}V_2(w,\theta) - 2\frac{1}{nh_d^k}V_3(w,\theta) + o\left(\frac{1}{n(\min\{h_n,h_d\})^k}\right).
\end{aligned}
$$

■

Lemma 4. *Let the assumptions of Theorem 4 be satisfied, $\hat{f}_{C|WZ}$ be as defined in* (5.7) *and Z_{ni} be as defined in the proof of Theorem 4. Then,*

$$
\sum_{i=1}^{n}\mathbb{E}\left|Z_{ni}/\sqrt{nVar(Z_{ni})}\right|^3 \to 0 \qquad as\ n\to\infty
$$

if $\alpha^3/(nh_n^k)\to 0$.

Proof. Note that

$$
\begin{aligned}
\sum_{i=1}^{n}\mathbb{E}\left|Z_{ni}/\sqrt{nVar(Z_{ni})}\right|^3 &= n(nVar(Z_{n1}))^{-\frac{3}{2}}\mathbb{E}|Z_{n1}|^3\\
&= n^{-\frac{1}{2}}(Var(Z_{n1}))^{-\frac{3}{2}}\mathbb{E}|Z_{n1}|^3 \qquad \text{(G.5)}
\end{aligned}
$$

and $\mathbb{E}|Z_{n1}|^3 = \mathcal{O}\left(\alpha^{-3/2}h_n^{-2k}\right)$ by Assumption 9 *(i)*. Moreover, by Assumption 9 *(ii)* there exists a constant $\kappa > 0$ such that $Var(Z_{n1}) > \kappa\alpha^{-2}h_n^{-k}$. Therefore,

$$\sum_{i=1}^{n} \mathbb{E}\left|Z_{ni}/\sqrt{nVar(Z_{ni})}\right|^3 = \mathcal{O}\left(\sqrt{\frac{1}{n\alpha^3 h_n^{4k}(Var(Z_{n1}))^3}}\right) = \mathcal{O}\left(\sqrt{\frac{\alpha^6 h_n^{3k}}{n\alpha^3 h_n^{4k}}}\right)$$

which converges to 0 if $\alpha^3/(nh_n^k) \to 0$.

■

Lemma 5. *Let the assumptions of Theorem 6 be satisfied. Then,*

$$(\alpha\tilde{P}_f + \tilde{P}_f T^* T\tilde{P}_f)f_{\theta|W}^{\alpha,L_f} = \tilde{P}_f T^* T\tilde{P}_f f_{\theta|W}^{\dagger c}$$

where $f_{\theta|W}^{\alpha,L_f}$ *is the solution of (5.9) with* $\widehat{f}_{C|WZ}$ *replaced by* $f_{C|WZ}$ *and* $\mathcal{F}_{\Theta|W}$ *replaced by* $L_f = \left\{h \in L^2_{\pi_\theta} : \left\langle f_{\theta|W}^{\dagger c} - f, h - f_{\theta|W}^{\dagger c}\right\rangle = 0\right\}$.

Proof. From (A.15) in the proof of Theorem 6 and because $f_{\theta|W}^{\alpha,L_f} \in L_f$ implies $(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}) \in \tilde{L}_f$, which in turn implies that $\tilde{P}_f(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}) = (f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c})$, we get: $\tilde{P}_f(T^*T + \alpha I)f_{\theta|W}^{\alpha,L_f} = \tilde{P}_f T^* f_{C|WZ}$ by using (A.15). Using these results, the following equivalences hold:

$$\begin{aligned}
\tilde{P}_f T^*T(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}) + \alpha\tilde{P}_f f_{\theta|W}^{\alpha,L_f} &= \tilde{P}_f T^* f_{C|WZ} - \tilde{P}_f T^*T f_{\theta|W}^{\dagger c} \\
\Leftrightarrow \quad \tilde{P}_f T^*T\tilde{P}_f(f_{\theta|W}^{\alpha,L_f} - f_{\theta|W}^{\dagger c}) + \alpha\tilde{P}_f f_{\theta|W}^{\alpha,L_f} &= \tilde{P}_f T^* f_{C|WZ} - \tilde{P}_f T^*T f_{\theta|W}^{\dagger c} \\
\Leftrightarrow \quad (\alpha\tilde{P}_f + \tilde{P}_f T^*T\tilde{P}_f)f_{\theta|W}^{\alpha,L_f} &= \tilde{P}_f T^* f_{C|WZ} + \tilde{P}_f T^*T(\tilde{P}_f - I)f_{\theta|W}^{\dagger c} \\
&= \tilde{P}_f T^*\left(T f_{\theta|W}^{\dagger c} + T(\tilde{P}_f - I)f_{\theta|W}^{\dagger c}\right) \\
&= \tilde{P}_f T^*T\tilde{P}_f f_{\theta|W}^{\dagger c}.
\end{aligned}$$

■

Lemma 6. *Suppose that assumptions 10 (i)-(iii) and (v) hold. Then, a solution to the minimization problem* (5.10) *exists.*

Proof. Problem (5.10) is numerically equivalent to the following procedure computed in two steps, where in the first step one computes, for each $\theta_1 \in \Theta_1$

$$m(\theta_1) = \min_{h \in \mathcal{F}_{\theta_2|W}} \left\{||T_{\theta_1}h - \hat{f}_{C|WZ}||^2 + \alpha||h||^2\right\} \tag{G.6}$$

and in the second step one computes:

$$\hat{\theta}_1 = \min_{\theta_1 \in \Theta_1} m(\theta_1). \tag{G.7}$$

A solution to (G.6) exists for every $\theta_1 \in \Theta_1$ since it is a convex problem. Moreover, under Assumptions 10 *(ii)-(iii)* and *(v)*, by Theorem 3 of Milgrom & Segal (2002) the value function $m(\cdot)$ is continuous. This together with compactness of Θ_1 implies the existence of a solution to (G.7). ■

Lemma 7. *The functional* $\xi(h) = \|h\|^2$ *defined on* $L^2_{\pi_\theta}$ *satisfies:*

$$\|h_1\|^2 - \|h_2\|^2 - \langle h_2, (h_1 - h_2)\rangle \geq c\|h_1 - h_2\|^2 \quad \textit{for any } 0 < c \leq 1.$$

Proof. Note that the Gâteaux derivative of $\|\cdot\|^2$ at $h_0 \in L^2_{\pi_\theta}$, denoted by $D(h_0)$ is equal to the linear functional $D(h_0) = \langle \cdot, h_0\rangle$ on $L^2_{\pi_\theta}$. Hence, for every $h_1, h_2 \in L^2_{\pi_\theta}$

$$D(h_0)(h_1 - h_2) = \langle D(h_1) - D(h_2), (h_1 - h_2)\rangle \geq c\|h_1 - h_2\|^2, \qquad \text{for any } 0 < c \leq 1. \tag{G.8}$$

Define $\varphi(t) = \|h_t\|^2$ where $h_t = th_1 + (1-t)h_2$, for $h_1, h_2 \in L^2_{\pi_\theta}$ and for every $t \in [0,1]$. Note that $h_t \in \mathcal{F}_{\theta|W}$ if $h_1, h_2 \in \mathcal{F}_{\theta|W}$. Moreover, $d\varphi(t)/dt = D(h_t)(h_1 - h_2)$. Then, for $0 \leq t' < t \leq 1$

$$\begin{aligned}
\frac{d\varphi(t)}{dt} - \frac{d\varphi(t')}{dt} &= \langle D(h_t) - D(h_{t'}), h_1 - h_2\rangle \\
&= \left\langle D(h_t) - D(h_{t'}), \frac{h_t - h_{t'}}{t - t'}\right\rangle \\
&\geq c\frac{\|h_t - h_{t'}\|^2}{t - t'}
\end{aligned}$$

where the second equality is due to the equality $h_t - h_{t'} = (t - t')(h_1 - h_2)$ and the last inequality is due to (G.8). Now, by setting $t' = 0$ we get $\frac{d\varphi(t)}{dt} - \frac{d\varphi(0)}{dt} \geq ct\|h_1 - h_2\|^2$. Therefore, by $\varphi(1) - \varphi(0) - \frac{d\varphi(0)}{dt} = \int_0^1 [\frac{d\varphi(t)}{dt} - \frac{d\varphi(0)}{dt}]dt \geq c\|h_1 - h_2\|^2$. By replacing $\varphi(1)$ with $\|h_1\|^2$ and $\varphi(0)$ with $\|h_2\|^2$ we get the result. ■

Lemma 8. *Under Assumption 10 (iv) and (vi) we have: (i)* $\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2 = O_p(\alpha^{-1}\delta_n)$ *for* $\delta_n = o(1)$*; (ii) if* $\delta_n = O(\alpha)$*, then there exists an* M_0 *such that:*

$$P\left(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 > M_0\right) \to 0 \qquad \textit{as } n \to \infty.$$

Proof. By definition of $\hat{g}$ and since $\widehat{Q}_n(\hat{g}) \geq 0$: $\alpha\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq \widehat{Q}_n(\hat{g}) + \alpha\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2$. This implies that

$$\begin{aligned}
\alpha\left(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2\right) &\leq \widehat{Q}_n(\hat{g}) + \alpha\left(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2\right) \\
&\leq \widehat{Q}_n(g^0) + \alpha\left(\|f^0_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2\right) \\
&= Q(g^0) + |\widehat{Q}_n(g^0) - Q(g^0)| + \alpha\left(\|f^0_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2\right) \\
&= O_p(\delta_n)
\end{aligned}$$

by Assumption 10 *(iv)* and *(vi)*, where $\delta_n = o(1)$ and the second inequality follows from the fact that $\check{f}^{\alpha,c}_{\theta_2|W}$ is the minimizer of the criterion (and hence, $\widehat{Q}_n(\hat{g}) + \alpha\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq \widehat{Q}_n(g^0) + \alpha\|f^0_{\theta_2|W}\|^2$).

This shows *(i)*. To show *(ii)* we use result *(i)* and observe that

$$\begin{aligned} P\left(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2 > M\right) &= P\left(\alpha(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 - \|f^0_{\theta_2|W}\|^2) > \alpha M\right) \\ &= P\left(O_p(\delta_n) > \alpha M\right) \end{aligned}$$

which converges to zero for every $M > 0$ if $\delta_n = O(\alpha)$. Finally, because $\|f^0_{\theta_2|W}\|^2$ is bounded, we can choose a finite $M_0 > 0$ sufficiently large so that $P\left(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 > M_0\right) \to 0$ as $n \uparrow 0$. ■

Lemma 9. *Let $\mathcal{U}_w(g^0)$ denote an open neighborhood in $\mathcal{G}$ in the weak topology around g^0. Under Assumptions 7 and 10 (i),(iv)-(vi), and if $\delta_n = O(\alpha)$, where δ_n is as in Assumption 10 (vi), then:*

$$P(\hat{g} \notin \mathcal{U}_w(g^0)) \to 0 \qquad \textit{as } n \to \infty.$$

Proof. Since $P(A) \leq P(A \cap B) + P(B^c)$ for any measurable sets A and B and by recalling the notation $\hat{g} = (\hat{\theta}_1, \check{f}^{\alpha,c}_{\theta_2|W})$ then

$$P(\hat{g} \notin \mathcal{U}_w(g^0)) \leq P(\hat{g} \notin \mathcal{U}_w(g^0), \|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq M_0) + P(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 > M_0) \tag{G.9}$$

for some $M_0 > 0$ large. Lemma 8 shows that for any $\varepsilon > 0$ there exists an $M_0 = M_0(\varepsilon) > 0$ such that $P(\|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 > M_0) < \varepsilon$. So, we focus on the first probability in the right hand side.

$$\begin{aligned} & P(\hat{g} \notin \mathcal{U}_w(g^0), \|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq M_0) \qquad \text{(G.10)} \\ \leq\ & P\left(\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} [\widehat{Q}_n(g) + \alpha\|f\|^2] \leq \widehat{Q}_n(g^0) + \alpha\|f^0_{\theta_2|W}\|^2\right) \\ \leq\ & P\Big(\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} [Q(g) + \alpha\|f\|^2] - \sup_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} |\frac{1}{2}\widehat{Q}_n(g) - Q(g)| \\ & \quad \leq Q(g^0) + \alpha\|f^0_{\theta_2|W}\|^2 + |\widehat{Q}_n(g^0) - Q(g^0)|\Big). \end{aligned}$$

Note that $|\frac{1}{2}\widehat{Q}_n(g) - Q(g)| = \frac{1}{2}\|T_{\theta_1}h - \hat{f}_{C|WZ}\|^2 - ||T_{\theta_1}h - f_{C|WZ}||^2 \leq \|\hat{f}_{C|WZ} - f_{C|WZ}\|^2$ by using the inequality $\frac{a^2}{2} - b^2 \leq (a-b)^2$. So, by Assumption 7, $\sup_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} |\frac{1}{2}\widehat{Q}_n(g) - Q(g)| = \mathcal{O}_p(\eta_n)$ for some $\eta_n = o(1)$. Then, from (G.10) and because $|\widehat{Q}_n(g^0) - Q(g^0)| = O_p(\delta_n)$, it follows:

$$\begin{aligned} & P(\hat{g} \notin \mathcal{U}_w(g^0), \|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq M_0) \\ \leq\ & P\left(\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} [Q(g) + \alpha\|f\|^2] \leq \alpha\|f^0_{\theta_2|W}\|^2 + O_p(\eta_n) + O_p(\delta_n)\right) \\ \leq\ & P\left(\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} Q(g) + \inf_{\|f\|^2 \leq M_0} \alpha\|f\|^2 \leq \alpha\|f^0_{\theta_2|W}\|^2 + O_p(\eta_n) + O_p(\delta_n)\right) \\ \leq\ & P\left(\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} Q(g) \leq O_p(\max\{\alpha, \eta_n, \delta_n\})\right). \end{aligned}$$

Note that the set $\mathcal{U}_w^c(g^0)$ is closed and $\Theta_1 \times \{f \in \mathcal{F}_{\theta_2|W}; \|f\|^2 \leq M_0\}$ is closed and bounded. Thus, under assumption 10 *(v)*, Q is continuous which implies that there exists a $g^* \in \{g \in \mathcal{U}_w(g^0)^c; \|f\|^2 \leq M_0\}$ such that $\inf_{g \notin \mathcal{U}_w(g^0); \|f\|^2 \leq M_0} Q(g) = Q(g^*)$. Moreover, by Assumption 10 *(iv)* it must be $Q(g^*) > 0$. If this was not the case, then we would have $g^* = g^0$, but this is a contradiction of the fact that $g^* \in \mathcal{U}_w^c(g^0)$.
Because $O_p(\max\{\alpha, \eta_n, \delta_n\}) \to 0$ as $n \to \infty$ and $Q(g^*) > 0$, we conclude that

$$P(\hat{g} \notin \mathcal{U}_w(g^0), \|\check{f}^{\alpha,c}_{\theta_2|W}\|^2 \leq M_0) \to 0$$

which in turn implies that $P(\hat{g} \notin \mathcal{U}_w(g^0)) \to 0$ as $n \to \infty$. ■

References

Heckman, J. & B. Singer (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52(2), 271–320.

Milgrom, P. & I. Segal (2002) Envelope theorems for arbitrary choice sets. *Econometrica* 70(2), 583–601.

Teicher, H. (1961) Identifiability of mixtures. *Annals of Mathematical Statistics* 32(1), 244–248.