Semiparametric Estimation of Random Coefficients in Structural Economic Models

Stefan Hoderlein[1] Lars Nesheim[2] Anna Simoni[3]

[1]Boston College

[2]CeMMAP, UCL and IFS

[3]Università Bocconi

MIT May 28, 2011.

Motivation

Heterogeneous population is characterized by the following first order condition

$$\partial_c u(c_t, \gamma) = \beta \mathbb{E}\left[R_{t+1} \partial_a v_t(W_{t+1}, Z_{t+1}, \theta) | W_t, Z_t\right] \quad (1)$$

where

- c_t is consumption,
- R_{t+1} is the interest rate,
- (W_t, Z_t) are state variables,
- $\theta = (\beta, \gamma)$ is a finite dimensional parameter vector, and
- (u, v_t) are **known** functions.
  - e.g., a CRRA utility function and the corresponding value function.

Question

- Solution of (1) implicitly defines **consumption function** $c = \varphi(w, z, \theta)$.
- Suppose φ is **known**.
- Suppose data on $(C_t | W_t, Z_t)$ are generated from composition of φ and an **unknown distribution** $f_{\theta|W}$.
- **Question.** Given data and knowledge of φ, can one **identify and estimate** $f_{\theta|W}$ nonparametrically?
  - **Knowledge** of φ and $f_{\theta|W}$ necessary to predict **distribution of impacts** of counterfactual changes in interest rates, income tax, pension and savings policy, etc.
  - Economic **theory provides information/structure** on φ; does not have much power to constrain $f_{\theta|W}$.
  - In general, economic logic implies θ and W are correlated.

Answer

- **Answer:** We show that:
  - For Z_t exogenous,

$$f_{C_t|W_t Z_t} = T_g f_{\theta|W}$$

    where T_g is an integral operator.
  - The **identified set** is the **set of solutions** of the previous equation that are densities.
  - **Estimation** can be based on regularization of the **pseudo-inverse** of T_g and computation of **null space** of T_g.
- In Euler equation case, can be much more flexible about φ.
- Our approach applies to **general non-separable structural models** of the form

$$\Psi\left(C, W, Z, \theta, \varepsilon\right) = 0.$$

Literature I

- *Nonparametric Random Coefficient Models:* Beran, Hall, Feuerverger (1994), Hoderlein, Klemela, Mammen (2004, 2009), Gautier and Kitamura (2010), Hoderlein (2010).
- *Nonparametric IV Models:* Florens (2002), Darolles, Florens, Renault (2002), Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, Kristensen (2007)....
- *Identification in Nonlinear Random Coefficient Models:* Bajari, Fox, Kim and Ryan (2009), Fox and Gandhi (2010)

Literature

- *Mixture Models:* Heckman and Singer (1984), Henry, Kitamura, Salanie (2010), Kasahara and Shimotsu (2009), Bonhomme (2010).
- *Parametric Consumption Models:* Deaton (1992), Alan and Browning (2009), Blundell, Browning and Meghir (1994), Browning and Lusardi (1996), Attanasio and Weber (2010), Gourinchas and Parker (2002)...

Contributions relative to literature I

Identification in Nonlinear Random Coefficient Models

- Provide general identification results using different assumptions (continuum of types vs. finite number).
- Provide formal statement of difficulty of identification making use of inverse problem literature.
- Introduce regularization bias to make estimation feasible and provide large sample theory.
- Make clear how results relate to economic features of the model and provide additional insights about source of identification.

Contributions relative to literature II

Parametric Consumption Models

- Most of literature allows either for no heterogeneity or only observed heterogeneity.
- We focus on quite flexible unobserved heterogeneity.
- Alan and Browning (2010): Nonparametric vs parametric.
- Provide results on identification.

Contributions relative to literature III

Relative to Nonparametric Random Coefficient Models:

- Extends work of Beran et al.(1994), Hoderlein et al.(2009), Gautier and Kitamura (2010) to general non-separable models.

Contributions relative to literature IV

Relative to Nonparametric IV Models:

- Very different objects of interest.
- Very close in terms of tools.

Contributions relative to literature V

Relative to Mixture Models:

- Very different models. Similarity: estimating equation

$$f_Y(y) = \int f_{Y|\theta}(y;\theta) f_\theta(\theta) d\theta.$$

- Heckman and Singer (1984), $f_{Y|\theta}(y;\theta) = f_{Y|\theta}(y;\theta,\sigma)$ parametric with finite parameter of interest σ, f_θ nuisance parameter.
- Henry, Kitamura, Salanie (2010), Bonhomme (2010) same objective as HS, nonparametric extension, place finite mixture structure on f_θ.
- We: f_θ parameter of interest, structure *e.g.* through CRRA model on $f_{Y|\theta}$.

General model

- **Assumption 1. (Structural Model).** The random variables $(C, W, Z, \theta, \varepsilon)$ satisfy

$$\Phi(C, W, Z, \theta, \varepsilon) = 0 \text{ almost surely} \tag{2}$$

  where Φ is a Borel measureable function. In addition, equation (2) has a unique solution in C implicitly defining the Borel measureable **consumption function**

$$C = \varphi(W, Z, \theta, \varepsilon).$$

- C is an outcome variable; $C \in \mathbb{R}$, observed.
- W are endogenous variables; $W \in \mathbb{R}^k$, observed.
- Z are exogenous variables; $Z \in \mathbb{R}^l$, observed.
- θ are random parameters; $\theta \in \mathbb{R}^d$, unobserved.
- ε is a random scalar, not of primary interest; $\varepsilon \in \mathbb{R}$ unobserved:
  - measurement error
  - unobserved state variable.

Euler equation example

- In the Euler equation example,
  - C is consumption,
  - W is assets and lagged income,
  - Z is current labor income,
  - ε is private information about future income, and
  - θ are parameters that represent heterogeneity in preferences or beliefs.

Differentiability

- **Assumption 2. (Differentiability)**. Ψ is $\mathbf{C}^1$ in a neighborhood of the set of solutions of (2) and

$$\begin{aligned} \partial_c \Psi(c, w, z, \theta, \varepsilon) &\neq 0 \\ \partial_\varepsilon \Psi(c, w, z, \theta, \varepsilon) &\neq 0 \end{aligned}$$

almost everywhere on the solution set of (2).

Dependence conditions

- **Assumption 3**.(Distribution of ε). The variable ε has a known continuous distribution conditional on (θ, W, Z) with Radon-Nikodym derivative $f_{\varepsilon|\theta WZ}$.
- **Assumption 4**. (Conditional independence of Z). The variables $(C, Z, \theta|W)$ have a joint continuous distribution and $Z \perp\!\!\!\perp \theta | W$.

Support conditions

- **Assumption 5.** The densities $f_{C|WZ}$ and $f_{\theta|W}$ are strictly positive and bounded on their supports for almost every (W, Z). The support of $f_{\theta|W}$ does not depend on W.

Example 1

- Finite horizon Euler equation with CARA utility.
- Given assets a_t, income z_t and a shock to permanent income ε_t , consumer chooses consumption c_t.
- Consumer's value function defined by

$$v_t(a_t, z_t, \varepsilon_t) = \max_{\{c_t\}} \left\{ \begin{array}{c} -\frac{e^{-\gamma c_t}}{\gamma} + \beta \mathbb{E}[v_{t+1}(a_{t+1}, z_{t+1}, \varepsilon_{t+1}, \theta) | z_t] \\ \text{subject to} \\ a_{t+1} = R(a_t - c_t) \\ z_{t+1} = z_t + \varepsilon_t + \nu_{t+1} \end{array} \right\}$$

where

- $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, $\nu_t \sim N\left(0, \sigma_\eta^2\right)$,
- $\varepsilon_t \perp\!\!\!\perp \nu_t$ and $(\varepsilon_t, \nu_t, z_0) \perp\!\!\!\perp (\theta, a_0)$, and
- $\theta = (\beta, \gamma)$.

Consumption function I

- Optimal consumption function takes the form:

$$c_t = \phi_{1t} a_t + \phi_{2t}(z_t + \varepsilon_t) + M_t(\gamma, \beta)$$

where

$$M_t(\beta, \gamma) = \frac{\phi_{3t}(\ln \beta + \ln R)}{\gamma} + \phi_{4t} + 0.5\phi_{5t}\gamma. \qquad (3)$$

- Trivial but illuminating example.
- a_t and $\theta = (\beta, \gamma)$ statistically dependent because a_t determined by past savings decisions.
  - Dependence changes with age.
- Income process is independent of preferences.
- Defining $W_t = (A_t, Z_{t-1})$, this implies

$$Z_t \perp\!\!\!\perp \theta | W_t.$$

Consumption function II

- Innovations to income Z_t move consumption around through known function φ. These movements are independent of θ.
  - In this example, due to linearity, this is not very helpful.
  - More generally, φ_t is not *additively separable* (non-normal disturbances, CRRA utility, stochastic interest rates).
- (β, γ) affect outcome only through single index $m = M_t(\beta, \gamma)$.
  - Joint distribution of $(\beta, \gamma)| W$ not point-identified but distribution of $M_t | W_t$ is.
  - Stochastic variation in interest rates, can point-identify joint distribution.
- Estimation method can be applied to a more general Euler equation model.
  - See Hoderlein, Nesheim and Simoni (2011).

Notation

- Let (π_θ, π_{cz}) be nonnegative weighting functions on **spt** (Θ) and **spt** $(\mathcal{C} \times \mathcal{Z})$ respectively.
- Consider the spaces

$$L^2_{\pi_\theta} = \left\{ h : \int_\Theta h(\theta, w)^2 \pi_\theta d\theta < \infty, \ P^W - a.e. \right\}$$

and

$$L^2_{\pi_{cz}} = \left\{ \psi : \int_{\mathcal{C} \times \mathcal{Z}} \psi(c, z, w)^2 \pi_{cz} dcdz < \infty, \ P^W - a.e. \right\}.$$

- Let $\mathcal{F}_{\theta|W} \subset L^2_{\pi_\theta}$ and $\mathcal{F}_{C|WZ} \subset L^2_{\pi_{cz}}$ be the subsets of densities on Θ and conditional densities on $\mathcal{C} \times \mathcal{Z}$ that are **strictly positive** and **bounded** on their supports.

Characterization of $f_{\theta|W}$

Theorem (1.i.)

Let $f_{C|WZ} \in \mathcal{F}_{C|WZ} \subset L^2_{\pi_{cz}}$. Under Assumptions 1-5, $f_{\theta|W}$ is a solution of the **nonlinear problem**

$$f_{C|WZ} = T_g f_{\theta|W} \text{ subject to } f_{\theta|W} \in \mathcal{F}_{C|WZ} \quad P^W - a.s. \tag{4}$$

where $T_g : L^2_{\pi_\theta} \to L^2_{\pi_{cz}}$ is defined for all $h \in L^2_{\pi_\theta}$ as

$$\left(T_g h\right)(c, w, z) = \int_\theta \frac{f_{C|WZ\theta}(c, w, z, \theta)}{\pi_\theta} h(\theta, w)\pi_\theta d\theta. \tag{5}$$

- Kernel of operator is

$$\begin{aligned} g(c, w, z, \theta) &= \frac{f_{C|WZ\theta}(c, w, z, \theta)}{\pi_\theta} \\ &= \frac{\sum_{i=1}^{s} f_{\varepsilon|\theta WZ}\left(\varphi_i^{-1}, \theta, w, z\right) \left| \frac{\partial_c \Psi(c,w,z,\theta,\varphi_i^{-1})}{\partial_\varepsilon \Psi(c,w,z,\theta,\varphi_i^{-1})} \right|}{\pi_\theta} \mathbf{1}_{\mathbf{spt}_{C|WZ\theta}} \end{aligned}$$

Comments

- Equation (5) is classical *mixture of probability densities* or *Fredholm integral of the first kind*.
- If $f_{C|WZ\theta} \equiv f(c-\theta, w, z)$ we recover the *convolution formula*.
- If F_ε is degenerate with a point mass only at $\varepsilon = 0$, can define the operator as

$$(T_g h)(c, z, w) = \int_\Theta 1\{\varphi(w, z, \theta) \leq c\} h(\theta, W) d\theta.$$

- Simple inversion of T_g does not work since $\mathcal{R}(T_g)$ is non-closed and in general $\hat{f}_{C|WZ} \notin \mathcal{R}(T_g)$.
- Pseudo-inverse $T_g^\dagger$ is also unbounded.

Hilbert-Schmidt

- **Assumption 6.** It is possible to choose π_θ and π_{cz} such that for P^W-a.e. w we have:

$$\int_{\mathcal{C}\times\mathcal{Z}} \int_{\Theta} \left[f_{C|WZ\theta}(c; w, z, \theta)\right]^2 \frac{\pi_{cz}}{\pi_\theta} dcdzd\theta < \infty.$$

- *Hilbert-Schmidt assumption* is sufficient to guarantee the compactness of T_g.
- Compact operator T_g has at most a countable number of singular values accumulating only at 0.
- Let $\left\{\lambda_j, \varphi_j, \psi_j\right\}_{j=1}^{\infty}$ be the SVD of T_g.
- How fast $\lambda_j \downarrow 0$ depends on smoothness of T_g.
  - This determined by smoothness of Ψ, distribution of ε and the support of (C, Z, θ).

Identification I

Theorem (1.ii.)

Under Assumptions 1-5, a solution of (4) exists since $f_{C|WZ} \in T_g \mathcal{F}_{\theta|W}$*. The identified set is*

$$
\begin{aligned}
\Lambda &= \{h \in \mathcal{F}_{\theta|W} : T_g h = f_{C|WZ},\ P^W \text{a.e.}\} \\
&= \left(f^{\dagger}_{\theta|W} \oplus \mathcal{N}(T_g)\right) \cap \mathcal{F}_{\theta|W} \\
&= \left\{ h \in L^2_{\pi_\theta} : \begin{array}{c} h = f^{\dagger}_{\theta|W} + \sum_{j \geq 1; \lambda_j = 0} z_j \varphi_j \text{ for some } \{z_j\} \\ h > 0 \\ \int_{\Theta} h = 1 \end{array} \right\}
\end{aligned}
$$

where $f^{\dagger}_{\theta|W}$ *is the solution of minimal norm of* $\left\| T_g h - f_{C|WZ} \right\|$ *and* φ_j *are the eigenfunctions of* $T_g^* T_g$ *corresponding to the zero eigenvalues* λ_j*.*

Identification II

Three situations are possible:

1. T_g one-to-one: $\mathcal{N}(T_g) = \{0\} \Rightarrow f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$.
2. T_g one-to-one on $\mathcal{F}_{\theta|W} \Rightarrow \left(f^{\dagger}_{\theta|W} \oplus \mathcal{N}(T_g)\right) \cap \mathcal{F}_{\theta|W} = \{f^{\dagger}_{\theta|W}\}$.
3. T_g not one-to-one on $\mathcal{F}_{\theta|W} \Rightarrow$ many possible values of $\{z_1, z_2, \ldots\}$.

Identification III

Theorem (2.)

Under Assumptions 1-5, the following are equivalent sufficient conditions for point identification of $f_{\theta|W}$.

1. The operator T_g is one-to-one. *i.e.* $\mathcal{N}(T_g) = \{0\}$.
2. The distribution of θ conditional on (C, W, Z) is complete.
3. $\forall (c, z) \in \mathcal{C} \times \mathcal{Z}$, $\frac{f_{C|WZ\theta}(c;w,z,\theta)}{\pi_\theta(\theta)}$ belongs to a complete subset $\mathcal{C}$ in $L^2_{\pi_\theta}$, $P^W - a.s.$
   - *i.e.* the only element in $L^2_{\pi_\theta}$ which is orthogonal to $\mathcal{C}$ is 0.
4. $\mathcal{R}(T_g^*)$ is dense in $L^2_{\pi_\theta}$, *i.e.* $\overline{\mathcal{R}(T_g^*)} = L^2_{\pi_\theta}$.
5. If Assumption 6 holds, the singular values of T_g are strictly positive.
   - *i.e.* $\lambda_j > 0$ for all j.

Discussion

- In this case, the inverse T_g^{-1} exists and

$$f_{\theta|W}(\theta; w) = T_g^{-1} f_{C|WZ}(c; w, z), \quad P^W - \text{a.e.}$$

- Properties of the operator T_g (like smoothness) are determined by smoothness properties of φ and f_ε.
- Identified set is smaller than the identified set of the equation $T_g h = f_{C|WZ}$.
- Point-identification requires $\dim(C, Z) \geq \dim(\theta)$.

Identification IV

- *Sufficient* condition for point-identification of $f_{\theta|W}$.
- **Lemma 3.1:** Assume that $f_{C|\theta WZ}(c, \theta, w, z)$ takes the form

$$\exp\left[\tau(c, w, z)^T m(\theta)\right] k(c, w, z) h(\theta)$$

  where $\mathbf{spt}\left(\tau\left(C, W, Z\right)\right) = \mathbb{R}^d$, $h(\cdot)$ is a positive function, and m is globally invertible. Then, $f_{\theta|W}$ is point-identified.

- Note exponential interactions between $m\left(\theta\right)$ and $\tau\left(c, w, z\right)$ and support condition.

Example 1: Identification I

- Operator equation is

$$f_{C|WZ}(c,w,z) = \int_{\Theta} \frac{\exp(-\frac{1}{2}(\frac{c-\phi_1 w - M(\beta,\gamma)-z}{\phi_2\sigma_\varepsilon})^2)}{\sqrt{2\pi\phi_2^2\sigma_\varepsilon^2}} f_{\beta\gamma|W}(\beta,\gamma,w)\,d\beta d\gamma. \tag{6}$$

- Joint density of (β, γ) **not point identified**.
- To see this, use

$$m = M(\beta, \gamma)$$

to make change of variables.

Example 1: Identification II

- **Change of variables** implies

$$f_{C|WZ}(c,w,z) = \int_M \int_\Gamma \frac{\exp\left(-\frac{1}{2}\left(\frac{c-\phi_1 w-m-z}{\phi_2\sigma_\varepsilon}\right)^2\right)}{\sqrt{2\pi\phi_2^2\sigma_\varepsilon^2}} \widetilde{f}_{M\gamma|W}(m,\gamma,w)\,dm\,d\gamma \qquad (7)$$

$$= \int_M \frac{\exp\left(-\frac{1}{2}\left(\frac{c-\phi_1 w-m-z}{\phi_2\sigma_\varepsilon}\right)^2\right)}{\sqrt{2\pi\phi_2^2\sigma_\varepsilon^2}} \widetilde{f}_{M|W}(m,w)\,dm$$

where

$$\widetilde{f}_{M\gamma|W}(m,\gamma,w) = f_{\beta\gamma|W}\left(M^{-1}(m,\gamma),\gamma\right)\left|\frac{\partial M^{-1}(m,\gamma)}{\partial m}\right|.$$

Example 1: Identification III

- Conditional on M, γ has no impact on consumption.
  - No interactions between (C, Z) and γ.
- Nevertheless, marginal density $\widetilde{f}_{M|W}$ is point-identified.
- To see this, use Lemma 3.1 framework and note

$$\begin{aligned}
\tau(c, w, z) &= \frac{c - \phi_1 w - z}{\phi_2 \sigma_\varepsilon} \\
h(\theta) &= \exp\left(-\frac{1}{2}\left(\frac{m(\theta)}{\phi_2 \sigma_\varepsilon}\right)^2\right) \\
k(c, w, z) &= \frac{\exp\left(-\frac{1}{2}\left(\frac{c - \phi_1 w - z}{\phi_2 \sigma_\varepsilon}\right)^2\right)}{\sqrt{2\pi \phi_2^2 \sigma_\varepsilon^2}}
\end{aligned}$$

Example 1: Identification IV

- **Identified set** of densities $f_{\beta\gamma|W}$ is the set

$$\left\{ h : h = \widetilde{f}_{M|W}\,\widetilde{f}_{\gamma|MW} \cdot \left| \frac{\partial M}{\partial \gamma} \right| \text{ for some } \widetilde{f}_{\gamma|MW} \in \mathcal{F}_{\gamma|MW} \right\}.$$

- Null space of operator is

$$N\left(T_g\right) = \left\{ h : h_1\left(m\right) \cdot \Delta h_2\left(\gamma \,|\, m\right) \right\}$$

  wheres h_1 solves (7) and Δh_2 is the difference between two arbitrary conditional densities.

- Data provide no restrictions on **conditional density;** other than support restrictions.
- Nevertheless, data do provide **some** restrictions on the **joint density**.
- See simulation results to follow.
- First, how do we estimate the model?

Estimation I

1. First estimate $\widehat{f}_{C|WZ}$.
2. Then solve constrained problem

$$\min_{\{h\}} \left\| T_g h - \widehat{f}_{C|WZ} \right\| + \alpha \left\| h \right\|_s^2$$

subject to

$$\begin{aligned} h &\geq 0 \\ \int_{\Theta} h(\theta, w)\, d\theta &= 1. \end{aligned}$$

   1. Paper contains results for case where $f^{\dagger}_{\theta|W} \notin \mathcal{F}_{\theta|W}$, (solve the constrained minimization numerically, no closed-form for the estimator and slower rate of convergence).
   2. Here, assume $f^{\dagger}_{\theta|W} \in \mathcal{F}_{\theta|W}$.
3. Ill-posed inverse problem. Regularization required because $T_g^{\dagger}$ is unbounded.

Estimation II

1. Two-step procedure to solve problem.
   1. Compute $\widehat{f}^{\alpha,s}_{\theta|W}$ to solve

$$\min_{\{h\}} \left\| T_g h - \widehat{f}_{C|WZ} \right\| + \alpha \|h\|_s^2 .$$

   2. Compute projection

$$\Pi_c \widehat{f}^{\alpha,s}_{\theta|W} = \max\left(0, \widehat{f}^{\alpha,s}_{\theta|W} - \frac{c}{\pi_\theta}\right)$$

   where c satisfies

$$\int_\Theta \Pi_c \widehat{f}^{\alpha,s}_{\theta|W} \, d\theta = 1.$$

      1. Algorithm by Gajek (1986).

Estimation of $f_{\theta|W}$

- **Assumption 7.** $\{c_i, w_i, z_i\}_{i=1}^N$ is an i.i.d. sample used to construct an estimator $\widehat{f}_{C|WZ}^N$ of $f_{C|WZ}$ such that

$$\lim_{N\to\infty} E\left(\left\|\widehat{f}_{C|WZ}^N - f_{C|WZ}\right\|^2\right) = 0.$$

1. First replace $f_{C|WZ}$ with a (nonparametric) estimator $\hat{f}_{C|WZ}$.
   - Kernel density estimator.
2. Second, compute regularized version of $T_g^\dagger$.
   - Tikhonov regularization ($\hat{f}_{\theta|W}^\alpha(\theta; w)$).
   - Tikhonov regularization in Hilbert scales (generalized Tikhonov).
3. Third, project onto space of densities.
4. Finally, compute eigenfunctions of null space.

- Alternatively, replace steps 2. and 3. with one-step constrained minimization.

Rate of convergence I

- **Assumption 8.** For some $\beta > 0$ and $0 < M < \infty$ the structural density $f^{\dagger}_{\theta|W}$ is an element of the β-regularity space $\Phi_\beta(M)$ defined as

$$\Phi_\beta(M) = \left\{ f \in \mathcal{N}(T_g)^{\perp}; \quad \sum_j \frac{< f, \varphi_j >^2}{\lambda_j^{2\beta}} < M \right\}.$$

- Smoothness condition. $f^{\dagger}_{\theta|W}$ is more smooth when β is larger.
- When $M = \infty$ then, $\Phi_\beta = \mathcal{R}[(T_g^* T_g)^{\frac{\beta}{2}}]$.

Rate of convergence II

Theorem

Given Assumptions 1-5 and 8, the MISE associated with $\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W}$ with $s = 0$ is

$$\mathbb{E}\left(\left\|\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}\right\|\right)^2 = \mathcal{O}\left(\alpha^{\beta\wedge 2} + \frac{1}{\alpha}\mathbb{E}\left(\left\|\hat{f}_{C|WZ} - f_{C|WZ}\right\|\right)^2\right).$$

Moreover, if $\alpha \asymp \left(\mathbb{E}\left(\left\|\hat{f}_{C|WZ} - f_{C|WZ}\right\|\right)^2\right)^{-\frac{1}{\beta\wedge 2+1}}$ *then,*

$$\mathbb{E}\left(\left\|\mathcal{P}_c \hat{f}^{\alpha}_{\theta|W} - f^{\dagger c}_{\theta|W}\right\|\right)^2 = \mathcal{O}\left(\left[\mathbb{E}\left(\left\|\hat{f}_{C|WZ} - f_{C|WZ}\right\|\right)^2\right]^{\frac{\beta\wedge 2}{\beta\wedge 2+1}}\right).$$

- For $s = 0$, there is no benefit from $\beta > 2$.
- In the paper, we present analysis of rates of convergence using regularization for $s > 0$ (Hilbert scales).
  - When $f^{\dagger}_{\theta|W}$ is highly smooth, these have faster rates of convergence.

Rates of convergence III

- Suppose $\widehat{f}_{C|WZ}$ is estimated as

$$\hat{f}_{C|WZ}(c;w,z) = \frac{\frac{1}{nh^{1+k+l}} \sum_{i=1}^{n} K_h(c_i - c, c) K_h(w_i - w, w) K_h(z_i - z, z)}{\frac{1}{nh^{k+l}} \sum_{l=1}^{n} K_h(w_l - w, w) K_h(z_l - z, z)}. \quad (8)$$

  where K_h is a (generalised) kernel of order r and h is a vector of bandwidths.
- Under mild regularity conditions on the kernel and the operator T_g, the optimal rate of the Tikhonov estimator is

$$\inf_{\alpha,h} \mathbb{E}||\hat{f}^{\alpha}_{\theta|W} - f^{\dagger}_{\theta|W}||^2 \asymp n^{-\frac{2\rho(\beta\wedge 2)}{(2\rho+k)(\beta\wedge 2+1)}}$$

  where
  - ρ is the number of derivatives in W of $f_{C|WZ}$.
  - Optimal values of α and h are: $h \asymp n^{-\frac{1}{2\rho+k}}$ and $\alpha \asymp n^{-\frac{2\rho}{(2\rho+k)(\beta\wedge 2+1)}}$.
- Curse of dimensionality only in the dimension of the endogenous variables W.

Pointwise asymptotic normality

Lemma (pointwise asymptotic normality): Let $\hat{f}^{\alpha}_{\theta|W}$ be the Tikhonov regularized estimator described above. If $\alpha \asymp n^{-\frac{2\epsilon\rho}{(2\rho+k)(\beta\wedge 2+1)}}$ for $\epsilon > 1$ and $h \asymp n^{-\frac{1}{2\rho+k}+\varepsilon_h}$, $\varepsilon_h < 0$, then (under mild assumptions) for P^W-a.e. W

$$\sqrt{nh^k}\frac{\hat{f}^{\alpha}_{\theta|W}(\theta,w) - f^{\dagger}_{\theta|W}(\theta,w)}{\Omega(\theta,w)} \Rightarrow \mathcal{N}(0,1)$$

where

$$\Omega(\theta,w) = \sum_{j=1}^{\infty}\frac{\lambda_j^2}{(\alpha+\lambda_j^2)^2}\Omega_1(j)\,\varphi_j^2 + 2\sum_{j<l}^{\infty}\frac{\lambda_j\lambda_l}{(\alpha+\lambda_j^2)(\alpha+\lambda_l^2)}\Omega_2(j,l)\,\varphi_j\varphi_l.$$

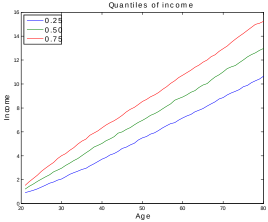
- Require α and h to converge faster than optimal to guarantee asymptotic bias of $\hat{f}^{\alpha}_{\theta|W}(\theta; w)$ is negligible.

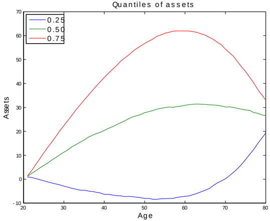
Estimation of Example 1

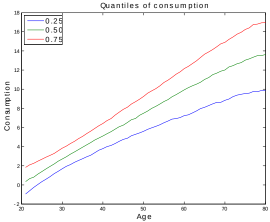
- Simulate data from CARA model.
- Estimate and plot the PDF of $M|W$.
- Estimate and plot the CDF of $M|W$.
- Display implications for identified set for (β, γ) .
  - Using $m = M(\beta, \gamma)$, plot level sets of M.
  - For each $u \in [0, 1]$, plot "quantile level sets" such as

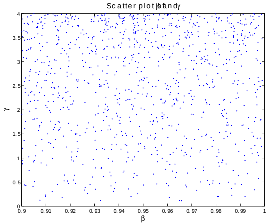
$$\Pr(M(\beta, \gamma) \leq m) = u.$$

Simulated data

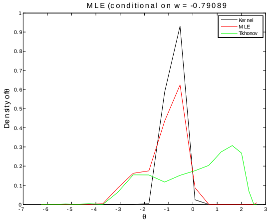


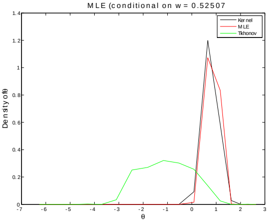


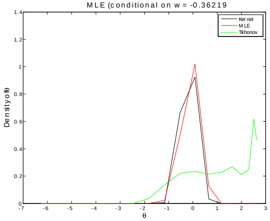


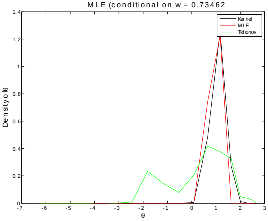


Example 1: Estimation I









Example 1: Estimation II