Dynastic Life-Cycle Discrete Choice Models: Estimation and Application

Limor Golan
Washington University in St. Louis

This lecture is based on:
Gayle, Golan and Soytas QE 2018, JME 2022

Econometric Society Summer School, DSE, December 15 2022, ANU.

Introduction

- Dynastic models:
  - Altruism has been recognized as an important factor underlying economic behavior of individuals.
  - Economic models with altruistic intergenerational links.
- It is important to model the intergenerational links explicitly.
  - Well-documented intergenerational persistence in education, income, and wealth.
  - Large literature showing that parental investment in children is important in the production function of human capital.
- Life-cycle component:
  - Sources of lifetime inequality are important to many economic and social policies.
  - Many decisions can be accounted for explicitly, otherwise would stay endogenous.

Framework for Estimation of Dynastic Models

- Many studies of social mobility.
- Growing literature understanding the mechanism.
- Goal: develop an estimator for dynastic models that allows us to estimate the model without solving it.
- Example for using it: Common measure IGC of child-parent income.
- Compare to the full solution method.
- The estimator can be used for discrete choice models and is extended to include a continuous choice (such as monetary investment in children).
- Application of altruistic unitary life-cycle model of household transmission of human capital across generations with endogenous fertility.

Estimation Issues

- The problem is a Discrete Markov Decision Process.
- Discrete Markov Decision Process:
  - In general have a representation in terms of valuation functions.
  - The full estimation (known as the Nested Fixed Point Algorithm) requires solution of the valuation functions for each parameter value in the search for optimal parameters (Miller 1984, Pakes 1986, Rust 1987).
  - Alternative proposed solution in the literature depends on the representation of the valuation function of the problem in terms of model primitives and estimated conditional choice probabilities (Hotz and Miller 1993).
- Intergenerational problem can be solved by a full estimation.
- However, a non-standard problem in Hotz and Miller technique.
  - Model combines both types of problem (finite and infinite horizons).
  - Does not satisfy the conditions for finite state dependency.
- Under stationarity assumption (over generations), a representation of the model valuation functions can be obtained depending only on model primitives and CCPs.
- The estimation can be done by a Pseudo Maximum Likelihood or a GMM estimator.

Generic Genderless Model

Environment

- The genderless individuals from each generation $g \in \{0, ...\infty\}$ live for $t = 0, ... T$ periods, where $t = 0$ is childhood and at $t = 1$ the individual becomes an adult.
- Adults in each generation derive utility from their own consumption, leisure, and from the utility of their adult offspring.
- Parents make decisions in each period about fertility, labor supply, time spent with children, and monetary transfers.
- No borrowing and savings.
- There can be intergenerational transfers; i.e. transfers of human capital, as in Loury (1981).

Model

Parental Choice

- Children consume and otherwise do nothing.
- Adults make discrete choices about:
  - Labor supply, h_t: individuals choose no work, part-time or full-time ($h_t \in (0,1,2)$).
  - Time spent with children, d_t: individuals choose none, low, and high ($d_t \in (0,1,2)$).
  - Birth, b_t: binary ($b_t \in (0,1)$).
- Let I_{kt} be an indicator for a particular choice k at age t.

Model

State Variables and Utility

- Define a vector, x, to include the time-invariant characteristics of education, skill, and race of the individual.
- Further define the vector z to include all past discrete choices as well as time-invariant characteristics:

$$z_t = (\{I_{k1}\}_{k=0}^{17}, ..., \{I_{kt-1}\}_{k=0}^{17}, x).$$

- An individual receives utility from consumption of a composite good, c_t denoted as $u_{2t}(c_t, z_t)$.
- Also different degrees of utility/disutility associated with choice k.
- The utility from I_{kt} is further decomposed in two additive components: a systematic component, denoted by $u_{1kt}(z_t)$, and an idiosyncratic component, denoted by ε_{kt}.
- $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$, the distribution function for ε_{kt} is assumed to be absolutely continuous with respect to the Lebesgue measure and has a continuously differentiable density.

Model

Expenditures and Budget Set

- For simplicity assume that the spending on children are proportional to an individual's current earnings and the number of children, but it depend on the state variables (parental education): $pc_{nt} = \alpha_{Nc}(z_t)(N_t + b_t)w_t(x, h_t)$.
- Budget constraint: $w_t(x, h_t) = c_t + \alpha_{Nc}(z_t)(N_t + b_t)w_t(x, h_t)$.
- We incorporate the budget constrain in $u_{2t}(c_t, z_t)$

Model

Intergenerational Transition Function

- $M(x'|z_{T+1})$ is a mapping of parental inputs and characteristics into a probability distribution of child's outcomes.
- z_{T+1} includes all parental choices and characteristics and contains information on the choices of time inputs and monetary inputs. Also contains information on all birth decisions - it captures the number of siblings and their ages.

Model

Life-cycle Dynastic model

- I^o_{kt} and c^o_t optimal choices; $u_{1kt}(z_t)$ utility associated with choices; $u_{2t}(c^o_t, z_t)$ utility from consumption; and an idiosyncratic component, denoted by ε_{kt}.
- Expected lifetime utility at time $t = 0$:

$$U_{gT}(x) = E_0\left[\sum_{t=0}^{T} \beta^t [\sum_{k=0}^{17} I^o_{kt}\{u_{1kt}(z_t) + \varepsilon_{kt}\} + u_{2t}(c^o_t, z_t)]|x\right]$$

- Altruistic preferences β, is the standard rate of time preference parameter.
- $\lambda N^{1-\nu}$ is the intergenerational discount factor, where N is the number of offspring an individual has over his lifetime.
- The total discounted expected lifetime utility of an adult in generation g:

$$U_g(x) = U_{gT}(x) + \beta^T \lambda E_0\left[N^{-\nu}\sum_{n=1}^{N} U_{g+1,n}(x'_n)|x\right],$$

where $U_{g+1,n}(x'_n)$ is the expected utility of child n ($n = 1, .., N$) with characteristics x'.

Optimal Discrete Choice

- The individual chooses the sequence of alternatives yielding the highest utility by following the decision rule $I(z_t, \varepsilon_t)$.
- We can write the value function of the problem, which represents the expected present discounted value of life time utility from following I^o, given z_t and ε_t, as:

$$
\begin{aligned}
V(z_{t+1}, \varepsilon_{t+1}) &= \max_I E_I \Big(\textstyle\sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{k=0}^{17} I_{kt'} [u_{kt'}(z_{t'}) + \varepsilon_{kt'}] \\
&\quad + \beta^{T-t'} \lambda N^{-\nu} \textstyle\sum_{n=1}^{N} U_{g+1,n}(x_n') | z_t, \varepsilon_t \Big) \qquad (1)
\end{aligned}
$$

By Bellman's principle of optimality, the value function can be defined recursively as:

$$
\begin{aligned}
V(z_t, \varepsilon_t) &= \max_I \Big[\textstyle\sum_{k=0}^{17} I_{kt} \left\{ u_{kt}(z_t) + \varepsilon_{kt} + \beta E(V(z_{t+1}, \varepsilon_{t+1}) | z_t, I_{kt} = 1) \right\} \Big] \\
&= \textstyle\sum_{k=0}^{17} I_{kt}^o (z_t, \varepsilon_t) [u_{kt}(z_t) + \varepsilon_{kt}] \\
&\quad + \beta \textstyle\sum_{z_{t+1}} \int V(z_{t+1}, \varepsilon_{t+1}) f(\varepsilon_{t+1}) d\varepsilon_{t+1} F(z_{t+1} | z_t, I_{kt}^o = 1) \qquad (2)
\end{aligned}
$$

Optimal Discrete Choice

- We define the conditional value function, $v_k(z_t)$, as the present discounted value (net of ε_t) of choosing k and behaving optimally from period $t = 1$ on:

$$v_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{z_{t+1}} V(z_{t+1}) F(z_{t+1} | z_t, I_{kt} = 1) \tag{3}$$

- The conditional value function is the key component to the conditional choice probabilities. Optimal decision rule at t can be rewritten using the individual's conditional value function:

$$I^o(z_t, \varepsilon_t) = \arg\max_{I} \textstyle\sum_{k=0}^{17} I_{kt}[v_k(z_t) + \varepsilon_{kt}] \tag{4}$$

- Model valuation function gives a implicit equation defining the *ex ante* value function as a function only the primitives of the model.

Optimal Discrete Choice-Ex-Ante Value Function and CCPs

- Defining the probability of choice k at age t by $p_k(z_t) = E[I^o_{kt} = 1|z_t]$
- Intergrating with respect to the density of the shocks, we can write: $V(z_t)$ as

$$\sum_{k=0}^{17} p_k(z_t)\left[u_{kt}(z_t) + E_\varepsilon[\varepsilon_{kt}|I_{kt} = 1, z_t] + \beta \sum_{z'} V(z')F(z'|z_t, I_{kt} = 1)\right] \quad (5)$$

This is standard, $V(z_t)$ is a function of the CCPs, the expected value of the preference shock, the per-period utility, the transition function, and the ex ante continuation value.

The conditional valuations function net of the shock is:

$$\upsilon_k(z_t) = u_{kt}(z_t) + \beta \sum_{z'} V(z')F(z'|z_t, I_{kt} = 1). \quad (6)$$

Optimal Discrete Choice-Ex-Ante Value Function and CCPs

can now be rewritten using the individual's optimal decision rule at t to solve

$$I^o(z_t, \varepsilon_t) = \arg\max_I \sum_{k=0}^{17} I_{kt}[v_k(z_t) + \varepsilon_{kt}]. \tag{7}$$

We can write the CCPs, $p_k(z_t)$ as a function of primitives and ex-ante valuation function

$$\int I^o(z_t, \varepsilon_t) f_\varepsilon(\varepsilon_t) d\varepsilon_t = \int \left[_{k \neq k'} 1\{v_k(z_t) - v_{k'}(z_t) \geq \varepsilon_{tk'} - \varepsilon_{kt}\}\right] f_\varepsilon(\varepsilon_t) d\varepsilon_t. \tag{8}$$

we will derives the implicit equation defining the ex ante value function as a function of only the primitives of the model.

Ex-ante Value Function Representation

- Now, we can express the components of the initial valuation function vector or matrix form.
- Note that there are no choices made and x is discrete with finte support

$$V_0 = \begin{bmatrix} V(1) \\ . \\ . \\ . \\ V(X) \end{bmatrix}, \quad U(k) = \begin{bmatrix} U_k(1) \\ . \\ . \\ . \\ U_k(X) \end{bmatrix}, \quad E(k) = \begin{bmatrix} e_k(p,1) \\ . \\ . \\ . \\ e_k(p,X) \end{bmatrix}, \; P(k) = \begin{bmatrix} p_k(1) \\ . \\ . \\ . \\ p_k(X) \end{bmatrix},$$

$$\iota_X = \begin{bmatrix} 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}_{Xx1}, \text{ and } M^o(k) = \begin{bmatrix} M_k^o(1|1) & \dots & M_k^o(X|1) \\ . & & \\ . & & \\ . & & \\ M_k^o(1|X) & \dots & M_k^o(X|X) \end{bmatrix}$$

Representation Theorem

- where I_X denotes the $X \times X$ identity matrix. Equation (12) is based on the dominant diagonal property,
- It implies that the matrix $I_X - \lambda \beta^T \sum_{k=0}^{17} \{P(k)\iota'_X\} \otimes (N_k^{-\nu} \otimes M^o(k))$ is invertible.
- The representation is obtained by combining known results from discrete choice estimation of stationary infinite-horizon problems with the finite horizon properties of the dynastic life-cycle model.
- See Aguirregabira and Mira (2002, 2007) and Pesendorfer and Schmidt-Dengler (2008) for the use and derivation of this inversion in context of stationary infinite horizon problems from discrete choice estimation of stationary infinite-horizon problems with the finite horizon properties of the dynastic life-cycle model.

Proposition

- There exists an alternative representation for the ex-ante conditional value function at time t which is a function of only the primitives of the problem and the conditional choice probability as:

$$\begin{aligned} \upsilon_k(z_t) &= u_{kt}(z_t) + \sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{s=0}^{17} \sum_{z_{t'}} p_s(z_{t'})[u_{st'}(z_{t'}) \\ &\quad + E_\varepsilon(\varepsilon_{st'}|I_{st'} = 1, z_{t'})] F_k^o(z_{t'}|z_t) + \lambda \beta^{T-t} {N_T}^{-\nu} \\ &\quad \cdot \sum_{n=1}^{N_T} \sum_x V(x) \sum_{s=0}^{K_T} \sum_{z_T} M_k^n(x'|z_T) p_s(z_T) F_k^o(z_T|z_t) \end{aligned} \tag{9}$$

where $V(x)$ is the corresponding element of the vector V_0 given as:

$$V_0 = [I_X - \lambda \beta^T N_T^{-\nu} \sum_{k=0}^{17} \{P(k)\iota_X'\} * M^o(k)]^{-1} \sum_{k=0}^{17} P(k)[U(k) + E(k)] \tag{10}$$

Representation Theorem

- Using these components the vector of the ex ante value functionbe expressed as

$$V_0 = \sum_{k=0}^{K} P(k) \otimes [U(k) + E(k) + \lambda\beta^T N_k^{-\nu} \otimes M^o(k) V_0] \tag{11}$$

- Rearranging the terms and solving for V_0, we obtain

$$V_0 = [I_X - \lambda\beta^T \sum_{k=0}^{17} \{P(k)\iota_X'\} \otimes (N_k^{-\nu} \otimes M^o(k))]^{-1} \sum_{k=0}^{17} P(k)[U(k) + E(k)], \tag{12}$$

A Generic Estimator of the Life-Cycle Dynastic Discrete Choice Model

- We parameterized
  - The period utility by a vector θ_2, $u_{kt}(z_t, \theta_2)$.
  - The period transition on the observed states by a vector θ_3, $F(z_t|z_{t-1}, I_{kT} = 1, \theta_3)$.
  - The intergenerational transitions on permanent characteristics by a vector θ_4, $M^n(x'|z_{T+1}, \theta_4)$.
  - The earnings function by a vector $\theta_5, w_t(x, h_t, \theta_5)$.
- Therefore the conditional value functions, decision rules, and choice probabilities now also depend on $\theta \equiv (\theta_2, \theta_3, \theta_4, \theta_5, \beta, \lambda, \nu)$.

Estimation - Representation 1

- *Non-standard problem, finite horizon within a generation, but an infinite horizon model.*
- **Step 1:** We estimate the CCP, transition, earnings functions necessary to compute the inversion.
- Estimate the CCPs, we denote this estimate by $\widehat{p_k(z_{dt1})}$. We also estimate θ_3, θ_4, and θ_5 which parameterize the transition and earnings functions $F(z_t|z_{t-1}, I_{kT} = 1, \theta_3)$, $M^n(x'|z_{T+1}, \theta_4)$ and $w_t(x, h_t, \theta_5)$ respectively in this step.
- **Step 2:** We derive representation of the ex-ante valuation function $V(x_0)$ in terms of CCP's, transition functions, per-period utility function parameters.
- **Step 3:** Can be estimated two ways:
  - Pseudo maximum likelihood (PML).
  - Generalized method of moment (GMM).

PML and GMM Estimators

- We can use a pseudo maximum likelihood method and not a pure maximum likelihood estimator because part of the likelihood function is concentrated out using the data. With D dynasties, the PML estimates of $\theta_0 = (\theta_2, \beta, \lambda, \nu)$ are obtained via:

$$\widehat{\theta}_{0PML} = \arg\max_{\theta_0} \left(\sum_{dt1=1}^{D} \sum_{t=0}^{T} \sum_{k=0}^{17} I_{dt1} \ln[p_k(z_{dt1}; \theta_0, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)] \right)$$

  where $p_k(z_{dt1}; \theta_0, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)$ is the CCP defined in terms of model primitives.

- Or the GMM estimate of θ_0 is obtained via:

$$\widehat{\theta}_{02SGMM} = \arg\min_{\theta_0} [1/D \sum_{d=1}^{D} \xi_d(\theta_0)]' \widehat{\Phi} [1/D \sum_{d=1}^{D} \xi_d(\theta_0)] \tag{13}$$

  where $\xi_{dt}(\theta_0)$ is the vector of moment conditions at t, these vectors are defined as $\xi_{dt}(\theta_0) = (\xi_{1dt}(\theta_0), \xi_{2dt}(\theta_0), ...\xi_{17dt}(\theta_0))'$.

Monte Carlo Study

Utility Function:

- The period utility function is linear.
- Agent chooses whether to invest or not $k_t \in \{0, 1\}$ in each period $t \in \{0, 1\}$.
- She gets the following utilities associated with each choice:

$$u(c_t, k_t, \varepsilon_t) = \begin{cases} c_t + \epsilon_t(0) & \text{if } k_t = 0 \\ (1-\theta)c_t + \epsilon_t(1) & \text{if } k_t = 1 \end{cases}$$

Intra-generation transition:

- Each agent starts the lifecycle with a particular consumption value $c_t \in (0.5, 0.6, 0.7, 0.8, 0.9)$.
- Transition from one state to another is by the transition matrix $F(c_1 \mid c_0, k_0)$:

c	$k_0 = 0$					$k_0 = 1$				
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.5	0.85	0.13	0.02	0	0	1	0	0	0	0
0.6	0.04	0.85	0.09	0.02	0	0.1	0.9	0	0	0
0.7	0.01	0.04	0.85	0.09	0.01	0.13	0.27	0.6	0	0
0.8	0	0.01	0.05	0.85	0.09	0.01	0.11	0.28	0.6	0
0.9	0	0	0	0	1	0	0.04	0.13	0.23	0.6

Inter-generational transition:

- Next generation's starting consumption value c_0' depends on the sum of the investment decisions in the lifecycle where $D \in (0, 1, 2)$.
- c_0' is determined by the intergenerational transition function $H(c_0' \mid D)$.

c_0' :	0.5	0.6	0.7	0.8	0.9	D
	1	0	0	0	0	0
	0	0.1	0.4	0.4	0.1	1
	0	0	0.04	0.06	0.9	2

- If the agent choses to invest 2 in the lifecycle, the probability that $c_0' = 0.9$ is 0.9.
- If the agent choses not to invest , $c_0' = 0.5$ for sure.

The estimated parameter values and their computation time. Pseudo Maximum Likelihood (PML) corresponds to the estimation conducted by the new estimator using PML and ML estimation is by the Nested Fixed Point (NFXP).

	Pseudo Maximum Likelihood sample size(n)				Nested Fixed Point (ML) sample size(n)			
$\theta = 0.25$	1,000	10,000	20,000	40,000	1,000	10,000	20,000	40,000
mean	0.24473	0.24935	0.24886	0.24881	0.22714	0.24571	0.23320	0.24477
stdev	0.04991	0.01328	0.00915	0.00668	0.04884	0.01354	0.02135	0.01019
bias	-0.00527	-0.00065	-0.00114	-0.00119	-0.02286	-0.00429	-0.01680	-0.00523
MSE	0.00249	0.00017	0.00008	0.00005	0.00288	0.00020	0.00073	0.00013
$\lambda = 0.8$								
mean	0.80425	0.79745	0.79797	0.79673	0.77538	0.78966	0.76934	0.78855
stdev	0.11241	0.03175	0.02157	0.01587	0.09211	0.03244	0.03656	0.02063
bias	0.00425	-0.00255	-0.00203	-0.00327	-0.02462	-0.01034	-0.03066	-0.01145
MSE	0.01253	0.00100	0.00046	0.00026	0.00901	0.00115	0.00226	0.00055
$\beta = 0.95$								
mean	0.94208	0.95245	0.95037	0.95136	0.93441	0.95227	0.94603	0.95027
stdev	0.06276	0.01893	0.01301	0.00934	0.05322	0.01983	0.01820	0.01236
bias	-0.00792	0.00245	0.00037	0.00136	-0.01559	0.00227	-0.00397	0.00027
MSE	0.00396	0.00036	0.00017	0.00009	0.00305	0.00039	0.00034	0.00015
Avg Comp Time[1]	0.65	2.88	6.06	12.60	347.6	376.4	467.5	509.8

[1]All of the simulations are conducted using the GAUSS programming language on 2 CPU 1.66 GHz, 3 GB RAM laptop computer. Unit of time is seconds.

Large Sample Properties

- Under certain regularity conditions,

$$\sqrt{N}\left(\theta^{(D)} - \theta^{o}\right) \Rightarrow N(0, \Sigma(\theta^{o}))$$

$$\begin{aligned}\Sigma(\theta^{o}) \quad = \quad & E\left[\nabla_{\theta}\bar{\xi}_{d}(Z)\Omega_{d}^{-1}\nabla_{\theta}\bar{\xi}_{d}(Z)'\right]^{-1} E\left[\nabla_{\theta}\bar{\xi}_{d}(Z)\Omega_{d}^{-1}\left\{\bar{\xi}_{d}(Z)+\Phi(Z)\right\}\left\{\bar{\xi}_{d}(Z)+\Phi(Z)\right\}'\Omega_{d}^{-1}\nabla_{\theta}\bar{\xi}_{d}(Z)'\right] \\ & \times E\left[\nabla_{\theta}\bar{\xi}_{d}(Z)\Omega_{d}^{-1}\nabla_{\theta}\bar{\xi}_{d}(Z)'\right]^{-1}.\end{aligned}$$

More on asymptotics

Appendix: Large Sample Properties

- Well known in the econometric literature that under certain regularity conditions, pre-estimation does not have any impact on the consistency of the parameters in the subsequent steps of a multistage estimation (Newey, 1984; Newey and McFadden, 1994; Newey, 1994).
- The asymptotic variance, however, is affected by the pre-estimation.
- In order to conduct inference in this type of estimation, one has to correct the asymptotic variance for the pre-estimation.
- The method used for correcting the variance in the final step of estimation depends on whether the pre-estimation parameters are of finite or infinite dimension.
- Unfortunately, our estimation strategy combines both finite- and infinite-dimensional parameters. Combining results from two sources (Newey, 1984; Newey and McFadden, 1994), however, allows us to derive the corrected asymptotic variance for our estimator.

Large Sample Properties

Following Newey (1984), we can write the sequential-moments conditions for the first- and third-step estimation as a set of joint moment conditions:

$$\bar{\xi}_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi) = [\xi_{dF}(Z_d, \theta_3), \xi_{dM}(Z, \theta_4), \xi_{dW}(Z, \theta_5), \xi_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi)]',$$

where $\xi_{dF}(Z_d, \theta_3)$ is the orthogonality condition from the estimation of the lifecycle transition function, $\xi_{dM}(Z, \theta_4)$ is the orthogonality condition from the estimation of the generation transition function, $\xi_{dW}(Z, \theta_5)$ is the orthogonality condition from the estimation of the earnings equation, and $\xi_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi)$ is the moment conditions from the second-step estimation defined in Equation (13).

Large Sample Properties Appendix

Regardless of the estimation method used to estimate θ_3, θ_4, and θ_5 they can always be expressed as moment conditions. Let $\theta = (\theta_0, \theta_3, \theta_4, \theta_5)'$, with the true value denoted by θ^o. Each element of infinite dimensional parameter, ψ, can be written as a conditional expectation. Redefine each element as $\psi^k(z^k) = f_{z^k}(z^k) E\left[\widetilde{I}_{dk} \mid z^k\right]$, where $\widetilde{I}_{dkt} = [1, I_{dkt}]'$ for the estimation of $p_k(z_{dt})$. Therefore, $\psi^{k(D)}(z^k) = \frac{1}{D}\sum_{d=1}^{D} \widetilde{I}_{dk} J_{\delta_N}(z^k - z_d^k)$. The conditions below ensure that $\psi^{(D)}$ is close enough to ψ^o for D large enough, in particular that $\sqrt{D}\left\|\psi^{(N)} - \psi^o\right\|^2$ converges to zero.

A3: *There is a version of $\psi^o(z)$ that is continuously differentiable of order κ, greater than the dimension of z and $\psi_1^o(z) = f_z(z)$ is bounded away from 0.*

A4: $\int J(u)\,\mathrm{d}u = 1$ *and for all* $j < \kappa$, $\int J(u)\left(\bigotimes_{s=1}^{j} u\right)\mathrm{d}u = 0$.

A5: *The bandwidth,* δ_D, *satisfies* $D\delta_D^{2\dim(z)}/(\ln(D))^2 \to \infty$ *and* $D\delta_D^{2\kappa} \to 0$.
A6*: There exists a* $\Psi(Z)$, $\epsilon > 0$, *such that*

$$\left\|\nabla_\theta \bar{\xi}_d(Z,\theta,\psi) - \nabla_\theta \bar{\xi}_d(Z,\theta^o,\psi^o)\right\| \leq \Psi(Z)\left[\|\theta-\theta^o\|^\epsilon + \|\psi-\psi^o\|^\epsilon\right]$$

and $E[\Psi(Z)] < \infty$.

A7: $\theta^{(D)} \to \theta^o$ *with* Θ^o *in the interior of its parameter space.*
A8: *(Boundedness)*
(i) Each element of $\bar{\xi}_d(Z, \theta, \psi)$ *is bounded almost surely:* $E[\|\bar{\xi}_d(Z, \theta, \psi)\|^2] < \infty$;
(ii) $p_{dkt} \in (0, 1)$, *for all* k.
(iii) $\xi_{dF}(Z_d, \theta_3)$, $\xi_{dM}(Z, \theta_4)$ *and* $\xi_{dW}(Z, \theta_5)$ *are continuously differentiable in* θ_3,θ_4, *and* θ_5 *respectively.*