

Fuzzy regression discontinuity

Het *fuzzy regression discontinuity*-model (FRD-model) is geformuleerd in twee stappen:

$$D_i = \alpha_0 + \alpha_1 f(S_i - \bar{S}) + \alpha_2 I[S_i \geq \bar{S}] + \alpha_3 X_i + u_i \tag{1}$$

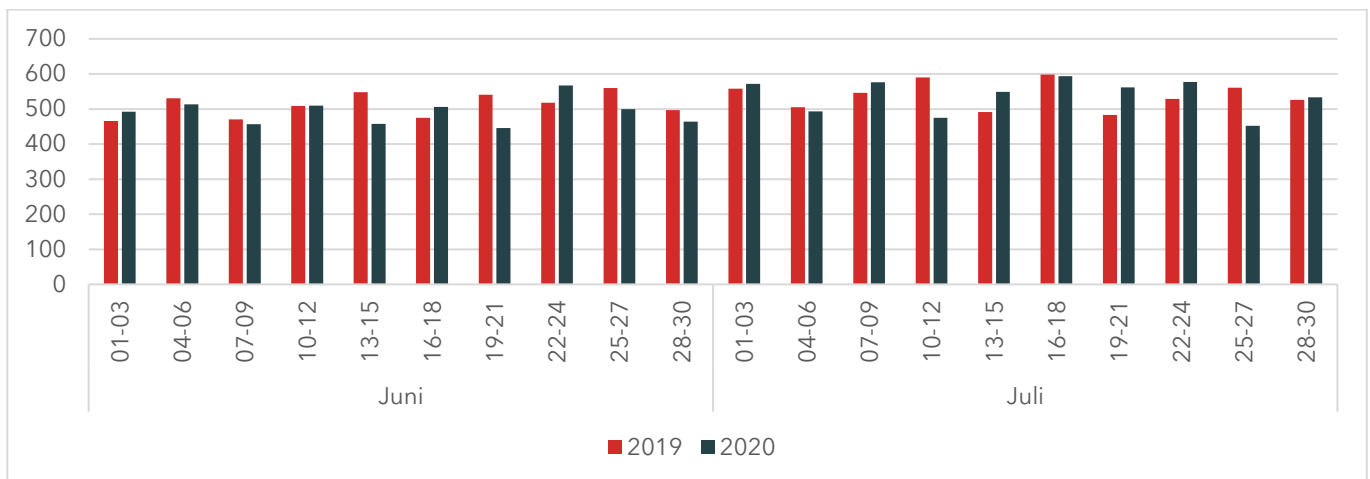
$$y_i = \beta_0 + \beta_1 \hat{D}_i + \beta_2 f(S_i - \bar{S}) + \beta_3 X_i + \varepsilon_i \tag{2}$$

Formule (1) vormt de eerste stap, waarin de endogene variabele D_i (opname verlof) wordt verklaard door de indicator $I[S_i \geq \bar{S}]$ die aangeeft of de geboorte vóór of na 1 juli 2020 heeft plaatsgevonden, door een functie $f(S_i - \bar{S})$ voor de periode tussen de geboortedag S_i en de grenswaarde (1 juli 2020) \bar{S} en door een aantal controlevariabelen X_i . Formule (2) gebruikt de geschatte waarden van \hat{D}_i om de uitkomstmaat y_i te verklaren, in dit geval het aandeel van de moeder in het huishoudensinkomen. De fouttermen zijn u_i en ε_i . De coëfficiënt β_1 geeft het effect van de opname van het aanvullend vaderschapsverlof weer.

Veronderstellingen

Een belangrijke aanname is dat huishoudens de indicator $I[S_i \geq \bar{S}]$ niet kunnen manipuleren. Met andere woorden: we nemen aan dat huishoudens niet in staat zijn de exacte geboortedag van het kind te bepalen om zo hun kans op vaderschapsverlof te vergroten. Enerzijds omdat er een grote variatie is rondom de uitrekendatum van vrouwen, en anderzijds omdat de uitrekendatum zelf ook maar beperkt beïnvloedbaar is: er zijn maar een klein aantal vruchtbare dagen in de menstruatiecyclus en er is een beperkte kans om zwanger te worden in deze dagen. Het is daarom aannemelijk dat vaders die net vóór, op of net na 1 juli 2020 een kind hebben gekregen *ex ante* vergelijkbaar zijn, maar door de invoering van de WIEG verschillen in de mate waarin ze toegang hebben tot aanvullend verlof.

Figuur 1 Verdeling van geboorten in onderzoekspopulatie in juni en juli 2019-2020



We onderbouwen deze veronderstelling door te kijken of er sprongen zijn in het aantal geboorten rondom 1 juli. Er zou immers in het geval van manipulatie een piek te zien moeten zijn in het aantal geboortes net na 1 juli 2020, in vergelijking met net vóór 1 juli 2020 en in vergelijking met eerdere jaren. Figuur 1 laat de distributie van geboortedagen zien rondom de grenswaarde (1 juli) in 2019 en 2020. Een *density test* van McCrary toont aan dat deze geboorteaantallen continu zijn rondom de grenswaarde ($p = 0,70$) (McCrary, 2008). Daarnaast blijkt uit tabel 1 dat de ouders die in juni 2020 een kind hebben gekregen, *ex ante* zeer vergelijkbaar zijn met de ouders die in juli dat jaar een kind hebben gekregen. Dit qua achtergrondkenmerken (leeftijd, opleiding) en qua arbeidsmarktpositie

(participatie, inkomen, sector, contractsoort). Immers, alle p -waarden van de t -testen zijn groter dan 0,05. Daaruit concluderen we dat er geen sprake is van manipulatie of zelfselectie in het recht op aanvullend verlof.

Tabel 1 Kenmerken van vaders en moeders in onderzoekspopulatie

	Vader			Moeder		
	Juni 2020	Juli 2020	p -waarde	Juni 2020	Juli 2020	p -waarde
Leeftijd	33,7	33,6	0,3165	30,9	30,9	0,8922
Opleidingsniveau						
Laag	9%	10%	0,0935	7%	7%	0,3859
Middelbaar	44%	45%	0,6854	40%	41%	0,3453
Hoog	36%	36%	0,4259	44%	44%	0,5076
Arbeidsmarktpositie						
Arbeidsparticipatie	97%	97%	0,0730	84%	84%	0,9122
Maandinkomen	€ 3.144	€ 3.126	0,6966	€ 1.983	€ 1.957	0,3058
Aandeel in huishoudensinkomen	64%	63%	0,6994	36%	37%	0,6994
Arbeidsuren per maand	156,4	155,4	0,2184	105,6	105,2	0,6754
Uurloon	€ 20,25	€ 20,17	0,7778	€ 18,52	€ 18,35	0,2243
Sector						
Private sector	75%	74%	0,5062	40%	40%	0,9349
Semipublieke sector	9%	9%	0,5709	30%	29%	0,5730
Publieke sector	13%	13%	0,5761	14%	15%	0,4677
Contractsoort						
Vast contract	90%	89%	0,0674	78%	78%	0,8766
Flexibel contract	7%	8%	0,3431	6%	6%	0,9210
Opname vaderschapsverlof						
Aandeel opname verlof	0%	38%		0%	0%	
Observaties	4.913	5.570		4.913	5.570	

Noot: De arbeidsmarktkenmerken zijn vastgesteld in augustus 2019, omdat de zwangerschapsperiode (waarin de arbeidsmarktkenmerken sterk veranderen) toen voor beide groepen nog niet was begonnen. De andere kenmerken zijn vastgesteld in de geboortemaand. Meerdere t -testen zijn uitgevoerd om te bepalen of de karakteristieke verschillen tussen ouders die in juni 2020 een kind hebben gekregen en ouders die in juli 2020 een kind hebben gekregen. De p -waarden van de t -testen zijn opgenomen in de tabel. Bij een p -waarde kleiner dan 0,05 is het verschil als significant beoordeeld.

Gevoeligheidsanalyses

We hebben een aantal gevoeligheidsanalyses uitgevoerd om de robuustheid van de bevindingen uit het basismodel te toetsen (tabel 2). Ten eerste hebben we gevarieerd in de controlevariabelen die we opnemen in het model. De resultaten zijn zeer vergelijkbaar als we geen, alleen achtergrondkenmerken (leeftijd, opleidingsniveau en regio) of een uitgebreide set aan controlevariabelen opnemen in het model (tabel 2, kolom 2). Dat versterkt ons vermoeden dat er sprake is van een experimentele setting. Ten tweede hebben we het model geschat met een kleinere bandbreedte van elf dagen rondom de grenswaarde van 1 juli, bepaald op basis van de optimale

bandbreedte test van Calonico et al. (2020). De effecten zijn zeer vergelijkbaar en tenderen zelfs nog iets sneller naar nul dan in het basismodel, maar dat kan ook komen door het kleinere aantal observaties. Ten derde gebruiken we verschillende specificaties voor de functie $f(S_i - \bar{S})$ in het model. Deze functie is in het basismodel gedefinieerd als de absolute afstand tussen de geboortedag en de grenswaarde (in dagen). Een veelgebruikte alternatieve specificatie in FRD-modellen is de polynoomfunctie van de afstand tot de grenswaarde. We specificeren daarom $f(S_i - \bar{S})$ als polynoom van de tweede graad, namelijk $(S_i - \bar{S}) + (S_i - \bar{S})^2$. We schatten ook een model zonder de afstand tot de grenswaarde op te nemen $f(S_i - \bar{S})$. De resultaten zijn echter zeer vergelijkbaar met het basismodel. Daarom concluderen we dat de resultaten robuust zijn voor andere modelspecificaties en data-selectie-procedures.

Tabel 2 Resultaten FRD-schattingen van opname aanvullend vaderschapsverlof: gevoeligheidsanalyse

	Aandeel van moeder in inkomen van het huishouden			
First stage				
Effect geboortemaand op opname verlof	0,38***	0,38***	0,40***	0,38***
<i>F</i> -statistiek	124	1.690	49	129
Second stage				
Effect verlof: 1 maand na geboorte	0,04***	0,04***	0,03***	0,04***
Effect verlof: 3 maanden na geboorte	0,02**	0,02	0,01	0,02**
Effect verlof: 5 maanden na geboorte	0,01	0,01	0,01	0,01
Effect verlof: 7 maanden na geboorte	-0,00	0,01	0,01	-0,00
Effect verlof: 9 maanden na geboorte	-0,00	0,00	0,00	-0,00
<i>Controlevariabelen</i>	Ja	Nee	Ja	Ja
<i>Bandbreedte rond 1 juli 2020</i>	1 maand	1 maand	11 dagen	1 maand
<i>Afstand tot grenswaarde in model</i>	Ja	Ja	Ja	Nee

Noot: */**/** Significant op respectievelijk tien, vijf- en eenprocentniveau. Controlevariabelen: afstand geboortedag tot grenswaarde, regio, leeftijd, opleidingsniveau, aantal kinderen en arbeidsmarktpositie (publieke/private sector, vast of flexibel contract, bedrijfsomvang) van zowel moeder als vader.

Tabel 3 geeft de schattingen weer van het FRD-model op het niveau van inkomen van zowel vaders (model 1) als moeders (model 2). Hieruit blijkt dat in de eerste drie maanden na de geboorte het inkomen van de vader significant afneemt. Dat is vermoedelijk het gevolg van de inkomensdaling tot zeventig procent bij vaders in verband met het opnemen van het verlof, waardoor de moeder automatisch een groter aandeel in het inkomen van het huishouden verdient. De effecten op het niveau van het inkomen zijn echter niet van blijvende aard, en verdwijnen na afloop van de verlofperiode, zoals te zien is in tabel 3.

Tabel 3. Resultaten FRD-schattingen van opname aanvullend vaderschapsverlof: niveau inkomen

	Niveau inkomen vader	Niveau inkomen moeder
First stage		
Effect geboortemaand op opname verlof	0,38***	0,38***
<i>F</i> -statistiek	210	195
Second stage		
Effect verlof: 1 maand na geboorte	-0,24***	0,03
Effect verlof: 3 maanden na geboorte	-0,13**	-0,10
Effect verlof: 5 maanden na geboorte	-0,20***	-0,22**
Effect verlof: 7 maanden na geboorte	-0,06	-0,08
Effect verlof: 9 maanden na geboorte	-0,05	-0,05
<i>Controlevariabelen</i>	Ja	Ja
<i>Bandbreedte rond 1 juli 2020</i>	1 maand	1 maand
<i>Afstand tot grenswaarde in model</i>	Ja	Ja

Noot: */**/** Significant op respectievelijk tien, vijf- en eenprocentniveau. Controlevariabelen: afstand geboortedag tot grenswaarde, regio, leeftijd, opleidingsniveau, aantal kinderen en arbeidsmarktpositie (publieke/private sector, vast of flexibel contract, bedrijfsomvang) van zowel moeder als vader.

Difference-in-differences event study

Het *difference-in-differences event study*-model is als volgt geformuleerd::

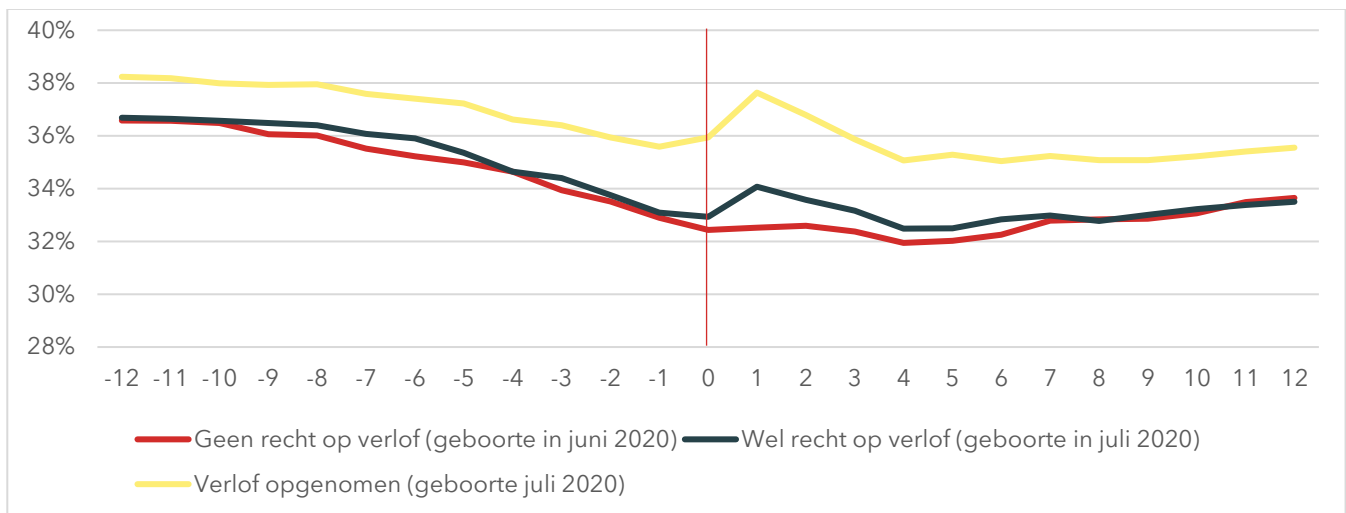
$$y_{it} = \beta_0 + \sum_{k=-12}^{-2} \rho_k (D_k \times T_i) + \sum_{k=0}^{12} \rho_k (D_k \times T_i) + \delta X_{it} + T_i + D_t + u_{it} \quad (3)$$

T_i is een dummyvariabele met de waarde 1 als het kind in juli 2020 is geboren. D_k is een dummyvariabele die gelijk is aan 1 als de periode gelijk is aan k . De periode $t = -1$ is uit het model gelaten om multicollineariteit te vermijden. X_{it} zijn controlevariabelen. Als het recht op verlof invloed heeft op het aandeel in het huishoudensinkomen dat de moeder verdient (y_{it}) in maand t na de geboorte, dan moet de coëfficiënt ρ_k significant verschillen van 0. Een vergelijking tussen coëfficiënten voor alle perioden na de geboorte $\rho_0, \dots, \rho_{T_1}$ maakt het mogelijk om de maanden na geboorte vast te stellen waarin het recht op verlof gevolgen heeft op het aandeel in het huishoudensinkomen dat de moeder verdient.

Veronderstelling

Een belangrijke aanname bij dit model is dat - in de afwezigheid van de WIEG - het verschil tussen behandel- en controlegroep constant is over tijd. Dat kunnen we onderbouwen door te kijken naar de mate waarin er sprake is van een gemeenschappelijke trend vóór invoering van de WIEG. Figuur 2 laat zien dat er over het algemeen sprake is van een gemeenschappelijke trend: in de twaalf maanden vóór het verlof veranderde het aandeel in het huishoudensinkomen dat de moeder verdient op vergelijkbare wijze in de controle- en behandelgroep.

Figuur 2 Trend in aandeel van moeder in inkomen van het huisouden, voor huishoudens die wel en geen recht op verlof hebben, en specifiek voor huishoudens die verlof hebben opgenomen



De uitkomsten van de *difference-in-differences event study* (zie figuur 1 in het artikel) laten wel zien dat in een aantal perioden een afwijking van de trend is waar te nemen, wat op zich niet verrassend is, gezien het grote aantal perioden dat we bekijken. Alleen in de derde maand vóór de geboorte gaat het om een relatief grote afwijking. We vermoeden dat het hier gaat om een corona-effect: de coronacrisis startte 3 maanden voor de geboorte voor de behandelgroep, maar nog niet volledig voor de controlegroep. Immers was het voor de behandelgroep drie maanden voor de geboorte april 2020, en voor de controlegroep maart 2020. Doordat vaders vaker werken in

sectoren die krompen door de coronacrisis dan moeders (die relatief vaak werken in bijvoorbeeld de zorg en andere cruciale beroepen), nam het huishoudaandeel van moeders naar verwachting toe in de coronacrisis. Zo blijkt uit onderzoek van Yerkes et al. (2020) dat 56 procent van de moeders en 34 procent van de vaders een cruciaal beroep heeft, en dat vaders in het begin van de coronacrisis meer zijn gaan zorgen voor de kinderen. Dat kan het positieve effect vóór de geboorte verklaren.

Gevoeligheidsanalyse

Tabel 4 geeft de resultaten van de gevoeligheidsanalyse weer. De coëfficiënten geven de verschillen weer in het aandeel huishoudinkomen van de moeder tussen de behandel- en controlegroep. Dit is het *average treatment effect*. We hebben gevarieerd in de controlevariabelen die we opnemen in het model. De resultaten zijn zeer vergelijkbaar als we geen, alleen achtergrondkenmerken (leeftijd, opleidingsniveau en regio) of een uitgebreide set aan controlevariabelen opnemen.

Tabel 4 Schattingen difference-in-differences event study met verschillende controlevariabelen

Maanden na geboorte	(1) geen controlevariabelen	(2) Basis controlevariabelen	(3) Uitgebreide controlevariabelen
-11	-0.0003	0.0000	0.0002
-10	0.0004	0.0004	0.0006
-9	0.0029**	0.0028*	0.0030**
-8	-0.0004	-0.0004	0.0003
-7	0.0012	0.0013	0.0020
-6	0.0014	0.0018	0.0028*
-5	-0.0025*	-0.0025*	-0.0011
-4	-0.0033***	-0.0035***	-0.0028**
-3	0.0050***	0.0050***	0.0057***
-2	-0.0013	-0.0014	-0.0004
-1	0.0000	0.0000	0.0017
0	0.0023*	0.0023*	0.0024**
1	0.0130***	0.0129***	0.0128***
2	0.0075***	0.0071***	0.0069***
3	0.0051**	0.0047**	0.0048**
4	0.0025	0.0020	0.0022
5	0.0014	0.0009	0.0012
6	0.0021	0.0015	0.0023
7	-0.0011	-0.0014	-0.0010
8	-0.0037	-0.0039	-0.0032
9	-0.0023	-0.0025	-0.0017
10	-0.0020	-0.0024	-0.0015
11	-0.0041	-0.0043	-0.0036
12	-0.0039	-0.0042	-0.0036
Observaties	10,483	10,482	10,482

Noot: */**/** Significant op respectievelijk tien, vijf- en eenprocentniveau. Controlevariabelen uitgebreid: afstand geboortedag tot grenswaarde, regio, leeftijd, opleidingsniveau, aantal kinderen en arbeidsmarktpositie (publieke/private sector, vast of flexibel contract, bedrijfsomvang) van zowel moeder als vader. Basis controlevariabelen: leeftijd, opleidingsniveau, regio