

**Academiejaar 2015 – 2016**

UNIVERSITEIT ANTWERPEN

FACULTEIT POLITIEKE EN SOCIALE WETENSCHAPPEN

FACULTEIT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

**Uptake en Effecten van Kinderopvang op Gezinsuitbreiding in Vlaanderen**

een Sociaal-Economische Analyse

Door Gert Thielemans

Masterproef voorgedragen tot het  
bekomen van de graad van:

Master in de Sociaal-Economische Wetenschappen



Promotor: Prof. dr. Karel Neels

Copromotor: Prof. dr. Sunčica Vujić  
Medebeoordelaar: Dhr. Pieter Vanpachtenbeke

**VOOR LIZA EN NELLES…**



**Academiejaar 2015 – 2016**

UNIVERSITEIT ANTWERPEN

FACULTEIT POLITIEKE EN SOCIALE WETENSCHAPPEN

FACULTEIT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

**Uptake en Effecten van Kinderopvang op Gezinsuitbreiding in Vlaanderen**

een Sociaal-Economische Analyse

Door Gert Thielemans

Masterproef voorgedragen tot het  
bekomen van de graad van:

Master in de Sociaal-Economische Wetenschappen

Promotor: Prof. dr. Karel Neels

Copromotor: Prof. dr. Sunčica Vujić  
Medebeoordelaar: Dhr. Pieter Vanpachtenbeke

## Dankwoord

Ik ben me ervan bewust dat het een onbegonnen zaak is om iedereen die deze thesis en bij uitbreiding de voltooiing van mijn masteropleiding mee mogelijk gemaakt heeft, te bedanken. Desondanks wil ik toch een bescheiden poging wagen.

Eerst en vooral dank aan mijn promotor, prof. Neels. Zonder zijn strenge en kritische opmerkingen was de kwaliteit van deze thesis nooit wat ze uiteindelijk geworden is. Zijn, vaak vlijmscherpe, maar altijd terechte commentaren zorgden zowel voor de haalbaarheid van het opzet als de kwaliteit van de uitvoering, zelfs al betekende dit dat er ergens een illusie aan diggelen geslagen moest worden. Dank ook aan mijn copromotor, prof. Vujić, die elke keer de scherven van mijn gebroken zelfvertrouwen vakkundig wist te lijmen wanneer dat gebeurde. Zonder de inzichten in en de interesse voor econometrie die ik van haar geleerd heb, was de vernieuwende aanpak die deze thesis biedt voor een bestaand probleem onmogelijk geweest.

Als 33-jarige student zijn er echter nog zoveel andere mensen die ik niet kan nalaten te bedanken. Bedankt aan mijn schoonouders, die me overtuigd hebben om de stap richting universiteit als dertiger toch te wagen. Bedankt ook vooral aan mijn eigen ouders, voor hun geduld. Het is omdat zij al die tijd hebben toegestaan dat ik mijn eigen weg zocht en me bleven steunen in mijn keuzes, dat er na veel omzwervingen eindelijk gebeurd is waar zij al minstens 10 jaar ongeduldig op zaten te wachten.

Dank ook aan mijn vrienden, die met hun wereldse kijk vaak een belangrijke inspiratiebron zijn. Zij helpen me waarschijnlijk meer dan ze zelf vermoeden. Bedankt dat ze luisteren wanneer ik vol overgave zat te ventileren over iets dat hen niet interesseert en vooral bedankt dat ze met één gevatte uitspraak de hele zaak konden relativeren wanneer het stressniveau weer eens te hoog lag.

Bedankt ook aan mijn kinderen, Liza en Nelles. Zelfs al zijn ze te jong om te weten waarom.

Maar vooral bedankt aan mijn vrouw, Aranka. Bedankt dat ik de kans heb gekregen om op een allesbehalve evident moment in ons leven tijdelijk afscheid te nemen van de arbeidsmarkt en me heb mogen concentreren op deze studies. Zij is en blijft het conditio sine qua non van deze hele onderneming.

## Abstract

Op basis van data uit de Belgische census van 2001, gecombineerd met registerdata en data van Kind en Gezin, werd in dit onderzoek het effect van het gebruik van formele kinderopvang voor een eerste kind op de kans op het krijgen van een tweede kind voor vrouwen in Vlaanderen onderzocht. De resultaten van probitregressies met instrumentele variabelen maken het mogelijk te controleren voor endogeniteit. Niet controleren voor endogeniteit blijkt te zorgen voor onderschatting van het effect, wat met het oog op gezinsbeleid tot verkeerde evaluaties en beslissingen zou kunnen leiden. Deze thesis is de eerste keer dat de techniek van instrumentele variabelen in dit soort setting gebruikt wordt en ondanks enkele kleine bemerkingen presteren ze voortreffelijk.

Uit de analyses blijkt dat vrouwen die gebruik maken van formele kinderopvang voor hun eerste kind jaarlijks 12% meer kans hebben op een tweede geboorte. Daarnaast werd bevestigd dat het effect groter is voor hoger opgeleide vrouwen. Voor vrouwen met een hogere kans op tewerkstelling werden slechts indicaties gevonden van een groter effect van het gebruik van formele kinderopvang op de kans op een tweede geboorte. Gelijktijdig gebruik van informele kinderopvang heeft dan weer geen invloed op het effect van formele kinderopvang. De resultaten zijn robuust voor modelspecificaties waarin rekening gehouden werd met regionale kenmerken van de arbeidsmarkt en de mate waarin een gemeente “kindvriendelijker” is.

De resultaten uit deze thesis een belangrijke mogelijke nieuwe richting voor het onderzoek naar de effecten van overheidsbeleid op vruchtbaarheid. Aanbieden en promoten van kinderopvang is een mogelijke oplossing voor landen die kampen met een zeer lage vruchtbaarheid. Verder onderzoek is echter nog nodig om het precieze effect voor alle geboortepariteiten te onderzoeken om zo uitspraken te kunnen doen over het effect van kinderopvang op het totale geboortecijfer.

## *Abstract (English)*

*Based on Belgian census data for 2001, combined with both register data and data from Kind & Gezin, the effect of the use of formal childcare for a first child on the probability of having a second child was investigated for women in Flanders. Results of probit-regressions with instrumental variables to control for possible endogeneity. Failure to do so is shown to results in a downward bias of the estimated effects which might lead to false policy evaluations and misinformed decision making. This thesis is the first of its kind to use probit regressions which allow for instrumental variables and despite some minor notes they perform remarkably well.*

*The analyses show that women who make use of formal childcare for their first child have an approximately 12% higher chance at a second birth per annum. Furthermore, the research confirms that the effect is higher for women with higher education. Only circumstantial evidence was found for the hypothesis that the effect of formal childcare is higher for women with better employment chances and no evidence was found that the use of informal childcare reduces the effects of formal childcare. Finally results proved robust for model specifications that control for regional characteristics of the lab our market and the “child-friendliness” of the municipality.*

*The results of this dissertation offer a new direction for the research on the effects of government policy on fertility. However, further research on the effect for all parity progressions is necessary in order to make conclusions of the overall impact of child care on fertility.*

***Keywords:*** *Fertility – Child Care – Endogeneity – Instrumental Variables*

Inhoudsopgave

[Lijst met figuren 2](#_Toc451926089)

[Lijst met tabellen 2](#_Toc451926090)

[Lijst met afkortingen 2](#_Toc451926091)

[Inleiding 3](#_Toc451926092)

[Deel 1: Literatuurstudie 5](#_Toc451926093)

[1. Theoretische literatuur 5](#_Toc451926094)

[1.1 New Home Economics – Gary Becker 5](#_Toc451926095)

[1.2 Richard Easterlin – endogene voorkeuren 7](#_Toc451926096)

[1.3 De rol van Kinderopvang 8](#_Toc451926097)

[2. Empirische literatuur 9](#_Toc451926098)

[3. Kinderopvang in Vlaanderen 13](#_Toc451926099)

[4. Hypothesen 15](#_Toc451926100)

[Deel 2: Empirische analyse 17](#_Toc451926101)

[5. Methodologie & Data 17](#_Toc451926102)

[5.1 Methodologie 17](#_Toc451926103)

[5.2 Data 20](#_Toc451926104)

[6. Resultaten van de analyses 23](#_Toc451926105)

[6.1 Beschrijvende analyse: het gebruik van kinderopvang 23](#_Toc451926106)

[6.2 Verklarende analyse: het effect van kinderopvang 26](#_Toc451926107)

[7. Discussie 31](#_Toc451926108)

[8. Beperkingen van het onderzoek 37](#_Toc451926109)

[8.1 Opzet 37](#_Toc451926110)

[8.2 Data 37](#_Toc451926111)

[9. Conclusies 39](#_Toc451926112)

[10. Toekomstig onderzoek 40](#_Toc451926113)

[Bibliografie 42](#_Toc451926114)

[Appendix 45](#_Toc451926115)

[A. Probit vs Logit vs LPM 45](#_Toc451926116)

[B. Validiteit en relevantie van de gebruikte instrumenten 46](#_Toc451926117)

[C. Eerste stadium IV-regressies (probit) 48](#_Toc451926118)

## Lijst met figuren

[Figuur 1: Gebruik van formele kinderopvang, naar opleidingsniveau 23](#_Toc451712877)

[Figuur 2: Gebruik van zowel formele als informele opvang naar opleidingsniveau 23](#_Toc451712878)

[Figuur 3: Gebruik van formele kinderopvang naar arbeidsmarktparticipatie 24](#_Toc451712879)

[Figuur 4: Gebruik van formele kinderopvang naar leeftijd 25](#_Toc451712880)

[Figuur 5: Gebruik van formele kinderopvang naar herkomst 25](#_Toc451712881)

[Figuur 6: Tweede geboorten en gebruik van formele kinderopvang 26](#_Toc451712882)

[Figuur 7: Marginaal effect van gebruik formele kinderopvang naar leeftijd bij geboorte van eerste kind 35](#_Toc451712883)

[Figuur 8: Marginaal effect van formele kinderopvang naar tijd sinds geboorte eerste kind (in jaren) 36](#_Toc451712884)

[Figuur 9: verdeling tewerkstellingskansen 38](#_Toc451712885)

[Figuur 10: Verdeling proportie gesubsidieerde opvang 38](#_Toc451712886)

## Lijst met tabellen

[Tabel 1: Overzicht empirische literatuur 14](file:///C:\Users\Gert%20Thielemans\OneDrive\school\Master\Thesis\preFinale%20versie.docx#_Toc451853012)

[Tabel 2: Beschrijving van de gebruikte variabelen 22](#_Toc451853013)

[Tabel 3: Resultaten van probitregressies 28](file:///C:\Users\Gert%20Thielemans\OneDrive\school\Master\Thesis\preFinale%20versie.docx#_Toc451853014)

[Tabel 4: Marginale effecten van het gebruik van formele kinderopvang bij gebruik = 0 30](#_Toc451853015)

[Tabel 5: Resultaten voor schattingen a.d.h.v. verschillende technieken en modelspecificaties 46](#_Toc451853016)

[Tabel 6: Robuustheidschecks voor gebruikte instrumenten a.d.h.v. lineaire regressie 46](#_Toc451853017)

[Tabel 7: Resultaten uit eerste stadium van IV-regressies, endogene variabele = gebruik van formele kinderopvang voor het eerste kind 48](#_Toc451853018)

## Lijst met afkortingen

Df Vrijheidsgraden

IV Instrumentele variabelen

LPM Linear probability model

SOEP Germans Socio-Economic Panel

TVC Totaal vruchtbaarheidscijfer

## Inleiding

Naar aanleiding van de spectaculair gestegen arbeidsmarktparticipatie van vrouwen sinds de jaren ’50 van de vorige eeuw en de gelijktijdige daling in de vruchtbaarheid tot onder het vervangingsniveau (Bratti, 2003; Neels, 2006), is een aanzienlijk corpus aan onderzoek ontstaan naar de oorzaken en gevolgen van deze evolutie. Een lage vruchtbaarheidsgraad leidt tot verschillende maatschappelijke problemen, waarvan de vergrijzing en de daaruit volgende betaalbaarheid van pensioenen en zorgkosten de meest zichtbare zijn (Chesnais, 1998). Aangezien aangehouden lage vruchtbaarheid op lange termijn negatieve gevolgen voor het arbeidsmarktaanbod en de overheidsinkomsten, hebben de meeste overheden ondertussen in meer of mindere mate gezinsbeleid ingevoerd dat er deels op gericht is om de vruchtbaarheid op peil te houden of te verhogen (McDonald, 2007).

In dit onderzoek wordt getracht een beter zicht te krijgen op de mogelijke rol die één van die maatregelen, namelijk kinderopvang, speelt in het vruchtbaarheidsvraagstuk.

Voorgaand onderzoek naar de rol van kinderopvang toont uiteenlopende resultaten naargelang het opzet (Gauthier, 2007). In cross-nationaal onderzoek werden eerder overwegend positieve effecten gevonden (Castles, 2003; Luci-Greulich, 2013; Lee & Lee, 2014). Deze resultaten zijn echter problematisch omdat als indicator voor vruchtbaarheid het *totale vruchtbaarheidscijfer* (TVC) gebruikt werd, zijnde het gemiddeld aantal kinderen per vrouw binnen een bepaalde periode (Kögel, 2006). Deze indicator negeert veranderingen in de voorkeuren op het vlak van vruchtbaarheid tussen vrouwen van verschillende geboortecohorten en timing-effecten zoals het uitstellen van een eerste geboorte en het tijdsinterval tussen twee kinderen (Sobotka & Lutz, 2009).

Een oplossing voor dit probleem is om niet te kijken naar het TVC, maar naar de individuele keuze voor gezinsvorming en –uitbreiding (Castles, 2003). In dit type onderzoek worden verbanden waargenomen op het volledige spectrum gaande van positief (Blau & Robins, 1991; Rindfuss et al., 2007, 2010; Haan & Wrohlich, 2009; Begall & Mills, 2011) over insignificant (Blau & Robins, 1989; Hank en Kreyenfeldt, 2003; Baizàn, 2009; Varga, 2014) tot verrassend negatief (Andersson, 2004; Krapf, 2009).

Een verklaring voor die verschillende resultaten is het bestaan van endogeniteitsproblemen, waarbij bijvoorbeeld beleid en vruchtbaarheid mekaar wederzijds beïnvloeden of de geobserveerde verbanden mogelijk (deels) toegeschreven kunnen worden aan de samenhang tussen de onderzochte variabelen en andere, niet-geobserveerde factoren (Neyer & Andersson, 2004; De Wachter & Neels, 2011).

In een bescheiden poging om een bijdrage te leveren aan het onderzoek naar overheidsbeleid in het algemeen en kinderopvang in het bijzonder en de rol hiervan in het vruchtbaarheidsvraagstuk, wordt in dit werkstuk het concept *vruchtbaarheid* (hier gezien als de *neiging* om kinderen te krijgen, in tegenstelling tot de biologische capaciteit ertoe) gedeconstrueerd en enkel gekeken naar de individuele keuze voor de pariteitsprogressie van een eerste naar een tweede kind. Op deze manier kan (dit onderdeel van) het vruchtbaarheidsprobleem bekeken worden in het licht van micro-economische (Becker, 1981; Easterlin, 1975) en sociologische (Friedman et al., 1994) rationele keuzetheorieën.

De onderzoeksvraag kan daarom specifieker gesteld worden als: welke rol speelt het gebruik van kinderopvang voor het eerste kind in de keuze voor een tweede kind?

In het eerste deel van deze thesis wordt een overzicht gegeven van de relevante bestaande theoretische en empirische literatuur over individuele vruchtbaarheidsbeslissingen en de rol die kinderopvang hierin mogelijk speelt. In het empirische gedeelte worden eerst de methode en de data beschreven die voor dit onderzoek gebruikt zijn. Aangezien deze data (Belgische censusdata uit 2001, gekoppeld aan registergegevens en data van Kind en Gezin) grotendeels cross-sectioneel van aard is, en mogelijk endogeniteit opduikt, wordt gebruik gemaakt van een probitmodel met instrumentele variabelen. Hiermee wijkt dit onderzoek methodologisch af van de analysetechnieken die courant in demografisch onderzoek gebruikt worden en kan het in die zin gezien worden als een nieuwe poging om inzichten uit demografisch onderzoek te koppelen aan econometrische analyse.

Het model wordt geschat voor twee verschillende specificaties waarbij de rol van kinderopvang in de individuele keuze voor een tweede kind wordt onderzocht en gecontroleerd voor a) individuele factoren b) omgevingsfactoren. De resultaten van deze analyses worden vervolgens op kritische wijze besproken. Na een uitgebreide discussie van de beperkingen van het onderzoek op het vlak van data, methodologie en veralgemeenbaarheid van de resultaten, worden tot slot enkele mogelijke beleidsimplicaties besproken en suggesties gedaan voor verder onderzoek.

De thesis is als volgt gestructureerd. In een eerste deel wordt een overzicht gegeven van de relevante theoretische en empirische literatuur en wordt een situatieschets gegeven van kinderopvang in Vlaanderen. In het tweede deel volgt een beschrijving van de gebruikte methode en de data. Daarna volgt een beschrijvende en verklarende analyse van de rol van kinderopvang. Na een uitgebreide bespreking van de resultaten worden enkele beperkingen van dit onderzoek besproken. Tot slot worden enkele conclusies en beleidsimplicaties geformuleerd en suggesties gedaan voor verder onderzoek.

# Deel 1: Literatuurstudie

## 1. Theoretische literatuur

De studie van gezinsvorming en gezinsuitbreiding is traditioneel gebaseerd op theorieën uit zowel de economische als de sociologische literatuur. De eerste heeft als voordeel elegante en eenvoudige modellen aan te bieden voor complexe problemen, maar krijgt als kritiek te reductionistisch te zijn (Friedman et al., 1994), te weinig rekening te houden met voorkeuren (Friedman et al., 1994; Gauthier, 2007) of te afhankelijk te zijn van assumpties die niet noodzakelijk aan de werkelijkheid voldoen (Easterlin, 1975; Friedman et al. 1994; Blau, 1998).

### 1.1 New Home Economics – Gary Becker

Een belangrijke poging om een economische verklaring te bieden voor vruchtbaarheid is het werk van Gary Becker uit 1960. Hoewel hij later grotendeels terugkwam op deze theorie (Yoram, 1982; Becker & Barro, 1986), blijven enkele inzichten nog steeds belangrijk voor de studie van gezinsvorming.

Eerst en vooral is er de assumptie dat gezinsvorming en –uitbreiding het gevolg zijn van rationele keuzes. Het is echter maar sinds de verspreiding en het grootschalige gebruik van contraceptie dat doorgedreven gezinsplanning mogelijk werd, hetgeen een uitbreiding van het aantal keuzemogelijkheden meebracht. Het zijn net die keuzemogelijkheden die in moderne samenlevingen onderzocht moeten worden (Easterlin, 1975).

De tweede assumptie die belangrijk gebleven is in de economische analyse van gezinsvorming, is de behandeling van kinderen als een duurzaam consumptiegoed. Door deze aanpak kan het nut van kinderen vergeleken worden met dat van andere goederen. De keuze voor een optimaal niveau van kinderen wordt hierdoor een maximalisatieprobleem, waarbij een zo groot mogelijk nut nagestreefd wordt, rekening houdend met een bepaalde budgetbeperking. Voorkeuren kunnen verschillen al naargelang religie, leeftijd, etniciteit enz., maar worden als exogeen beschouwd. Het model doet hier geen voorspellingen over, enkel over “economische” factoren (Becker, 1960).

Bovendien zijn kinderen een normaal goed, waardoor een hoger inkomen resulteert in een hogere *consumptie* van kinderen. Hierbij moet opgemerkt worden dat het “optimale niveau” van kinderen niet enkel duidt op een optimale kwantiteit van kinderen, maar ook rekening houdt met de kwaliteit. Duurdere kinderen, in termen van waarde die aan de opvoeding ervan gespendeerd wordt, zijn kinderen van hogere kwaliteit (Becker & Lewis, 1973). In die zin is een consumptie een trade-off tussen kwantiteit en kwaliteit en kan een lager aantal kinderen het gevolg zijn van een hogere consumptie in termen van kwaliteit.

Als gevolg kan de keuze voor een bepaald aantal kinderen verklaard worden door veranderingen in enerzijds het *inkomen* en anderzijds de *kostprijs van kinderen.* Naarmate het inkomen stijgt, wordt er meer uitgegeven wordt aan kinderen. Wanneer de elasticiteit van kwaliteit groter is dan die van kwantiteit zorgt inkomensstijging een stijging van beiden, maar de stijging in kwantiteit zal, zeker in moderne samenlevingen, eerder gering (of zelfs negatief) zijn ten opzichte van een toegenomen kwaliteit (Becker & Lewis, 1973).

De kostprijs van kinderen wordt door Becker gedefinieerd als de totale verwachte uitgaven inclusief de berekende kosten van diensten die door de ouders aangeboden worden minus de verwachte opbrengsten van kinderen inclusief de berekende waarde van de aangeleverde diensten door kinderen. In moderne samenlevingen wordt aangenomen dat kinderen een netto kost zijn (Becker, 1960). Voorts is het belangrijk om in te zien dat de verandering in kostprijs van kinderen een verandering is in de kost van kinderen van gelijke kwaliteit. Dit is belangrijk omdat een verandering in de uitgaven aan kinderen niet noodzakelijk een indicatie is van een gewijzigde prijs van kinderen (Gauthier, 2007).

Tot slot wordt aangenomen dat rationele actoren zo dicht mogelijk hun optimale gezinsgrootte proberen te benaderen. Aangezien kinderen enkel discreet geconsumeerd kunnen worden brengt dit mee dat de kwaliteit van kinderen verhoogt wanneer er, om welke reden dan ook, te weinig kinderen zijn in een gezin en omgekeerd dat er bespaard wordt op kwaliteit van kinderen wanneer het aantal te groot is (Becker, 1960).

In latere herwerkingen van zijn theorie, voornamelijk het in 1981 verschenen *a treatise on the family*, nam Becker zoals reeds aangehaald nadrukkelijk afstand van zijn oorspronkelijke theorie dat de grootte van een gezin sterk afhankelijk was van de kennis van contraceptie (Becker, 1981; Yoram, 1982). Nieuw is hier dat de voornaamste kostprijs van kinderen bepaald wordt door de *tijd* die vrouwen spenderen aan opvoeding. Hiervoor wordt uitgegaan van een genderspecifieke taakverdeling binnen het huishouden. De kost van kinderen wordt zo in grote mate afhankelijk van de opportuniteitskost van de afwezigheid van vrouwen op de arbeidsmarkt (Becker, 1981; De Wachter & Neels, 2011).

Onder de assumptie dat het grootbrengen van kinderen voornamelijk door de vrouw gebeurt en daarom onverenigbaar is met voltijdse participatie op de arbeidsmarkt, houdt dit in dat voor hoger opgeleide vrouwen, met een hoger potentieel loon, kinderen relatief duurder zijn in termen van opportuniteitskosten, met een negatief gevolg voor de totale consumptie van kinderen. Alles wat een invloed heeft op de kostprijs van kinderen of op het inkomen van ouders, zorgt zo voor *substitutie-* en *inkomenseffecten* met betrekking tot de kwaliteit en kwantiteit van kinderen, dus ook kinderopvang.

Samengevat zijn volgens deze klassieke economische interpretatie van gezinsvorming de vruchtbaarheidsbeslissingen (zowel kwantiteit als kwaliteit) van rationele actoren afhankelijk van enerzijds de kostprijs van kinderen en anderzijds het gezinsinkomen. Hierbij wordt de opportuniteitskost van ouderlijke zorg (voornamelijk door vrouwen) beschouwd als de belangrijkste kostprijs van kinderen.

### 1.2 Richard Easterlin – endogene voorkeuren

Hoewel hij een voorstander van de economische benadering van vruchtbaarheidsbeslissingen blijft, wordt de strakke assumptie van exogene voorkeuren door Easterlin (1975) nadrukkelijk verworpen. Als uitbreiding van de voorgaande theorie stelt hij voor om als determinanten van gezinsgrootte een onderscheid te maken tussen de vraag naar kinderen, de productie van kinderen en de kostprijs van vruchtbaarheidsregulering (contraceptie).

In navolging van Becker beschouwt Easterlin (1975) de vraag naar kinderen als een functie van het inkomen en de prijs van kinderen, maar hij voegt hier expliciet de endogene voorkeur aan toe, die tot dan toe volgens hem onderbelicht is gebleven in economische analyses. Zo zou de hogere prijselasticiteit van de kwaliteit van kinderen ten opzichte van de kwantiteit afhankelijk zijn van subjectieve voorkeuren. Een stijging in het inkomen zorgt bovendien voor een stijging van zowel de kwaliteit als de kwantiteit van kinderen, enkel bij veranderende inputprijzen is er substitutie tussen de twee.

De volledige nutsfunctie wordt gezien als endogeen aan de sociale context, waarbij een onderling afhankelijk voorkeursmechanisme speelt tussen individuele en ‘sociale’ voorkeuren dat aangeduid kan worden als *socialisering* (Easterlin et al., 1975). Vruchtbaarheidsbeslissingen worden genomen met als doel een nutsfunctie te maximaliseren, rekening houdend met beperkingen, maar met een focus op gewenste vruchtbaarheid. Vooral in moderne samenlevingen, waarin kennis en verspreiding van contraceptie ervoor zorgt dat perfecte gezinsplanning benaderd wordt (Bumpass & Westoff, 1970 volgens Bongaarts & Westoff, 2000), stijgt het belang van endogene voorkeuren naar ideale gezinsgrootte[[1]](#footnote-1).

Goldstein (2003) merkt op dat deze voorkeuren kunnen wijzigen doorheen de levenscyclus en andere auteurs benadrukken dat deze voorkeuren pariteitsafhankelijk kunnen zijn (Monnier, 1987; Yamaguchi & Ferguson, 1995). Ondanks de beperkingen omtrent de betrouwbaarheid van het meten van vruchtbaarheidsintenties, spelen voorkeuren een belangrijke rol in verschillende theorieën rond vruchtbaarheidsbeslissingen (Goldstein, 2003). Zo stellen Lesthaeghe & Surkyn (1988) dat socialisatie een belangrijke rol speelt in de voorkeur voor kinderen. Hierbij wordt in de eerste plaats gedacht aan cohortesocialisatie (een algemeen geldend waardeoordeel over het ideale aantal kinderen). Ook het theoretische model van Kohler (2001) beargumenteert dat de timing en snelheid waarmee vruchtbaarheidspatronen veranderen deels verklaard kunnen worden door sociale interactie. Hiermee wordt bedoeld dat gedrag (bijvoorbeeld het gebruik van contraceptie) en voorkeuren (bijvoorbeeld het gewenst aantal kinderen) beïnvloed worden door het gedrag en de attitudes van anderen (Kohler, 2001 volgens Bongaarts, 2002).

Regionale heterogeniteit in gemiddelde gezinsgrootte zou op die manier een invloed kunnen hebben op vruchtbaarheidsbeslissingen. Goldstein (2003) wijst er echter op dat de causaliteit volgens de sociologische theorie in beide richtingen kan lopen, waarbij normen zoals ideale gezinsgrootte een invloed kunnen hebben op vruchtbaarheidsuitkomsten, maar gerealiseerde gezinsgrootte van een vorige generatie eveneens een invloed kan hebben op de idealen van een huidige generatie.

### 1.3 De rol van Kinderopvang

De meest gangbare verklaring voor de mogelijke rol die kinderopvang speelt in de keuze voor kinderen is dat het de arbeidsmarktparticipatie van vrouwen vergemakkelijkt. Als tijd inderdaad de belangrijkste opportuniteitskost is voor vrouwen (Becker, 1981), zorgt het gebruik van kinderopvang ervoor dat deze kost gereduceerd wordt omdat het vrouwen in staat stelt tijd die normaal gespendeerd wordt aan zorg te gebruiken om een inkomen te verwerven op de arbeidsmarkt. Hierdoor stijgt het inkomen en de vraag naar kinderen. De precieze stijging van het aantal kinderen is zoals eerder aangehaald afhankelijk van de elasticiteiten van de vraag naar kwaliteit en kwantiteit van kinderen.

De prijs van kinderopvang bepaalt mee of vrouwen ervoor kiezen om te participeren op de arbeidsmarkt, aangezien de netto baten van participatie (o.a. stijging van het netto-gezinsinkomen) groter moeten zijn dan de netto-opportuniteitskosten van niet-participatie (o.a. het misgelopen loon). Hierdoor wordt de prijs van kinderopvang misschien problematisch voor laaggeschoolde vrouwen met (uitzicht op) een laagbetaalde job (Rindfuss, 2010). Het reservatieloon van vrouwen is daarnaast onder andere afhankelijk van verschillen in gezinsinkomen uit andere bronnen dan het (mogelijke) arbeidsinkomen van de vrouw (niet-arbeidsinkomen of het inkomen van de partner).

Wanneer we kinderopvang opsplitsen in formele vormen (créches, onthaalmoeders, au pair,…) en informele vormen (grootouders, familie, vrienden,…) en daarbij aannemen dat informele opvang gemiddeld de goedkopere oplossing is van de twee, spelen er substitutie-effecten die de keuze voor formele opvang mee bepalen.

Anderzijds spelen endogene voorkeuren zoals aangehaald in het werk van Easterlin (1975) evenzeer een rol in de keuze voor het gebruik van formele kinderopvang. Zo is het mogelijk dat culturele voorkeuren voor opvang door familieleden of discriminatie waardoor de toegang tot formele kinderopvang beperkt wordt ervoor zorgt dat bijvoorbeeld etnische achtergrond een rol speelt in de beslissing om gebruik te maken van (in)formele opvang (Uttal, 1999) .

Hieruit volgt dat het gebruik van kinderopvang samen hangt met de heterogeniteit onder vrouwen en gezinnen. Aangezien het gebruik van kinderopvang evenzeer een individuele keuze is, stelt zich hier een specifiek endogeniteitsprobleem, namelijk zelfselectie (cf. supra), wat de resultaten van analyses van het effect van kinderopvang op vruchtbaarheidsbeslissingen mogelijk vertekent.

## 2. Empirische literatuur

In dit onderdeel van de paper wordt een overzicht gegeven van de empirische literatuur die de rol van kinderopvang in individuele vruchtbaarheidsbeslissingen bestudeert. Daarbij wordt met bijzondere aandacht gekeken naar hoe de auteurs rekening houden met endogeniteit. Er volgt een uitgebreide beschrijving van dit probleem in het methodologische hoofdstuk, maar het is niettemin belangrijk met het oog op de kritische evaluatie van eerder onderzoek om kort de belangrijkste oorzaken van endogeniteit te vermelden.

Hill et al (2012) onderscheiden drie belangrijke oorzaken van endogeniteit: vertekening door niet-geobserveerde (of niet observeerbare) kenmerken die zowel correleren met de afhankelijke als de onafhankelijke variabele (*omitted variable bias*), vertekening door meetfouten (*measurement error*) en vertekening door simultaneïteit of omgekeerde causaliteit (*reverse causality / simultaneity bias*). Onderzoek dat nalaat te controleren voor dergelijke problemen geeft vertekende resultaten. Een belangrijk gevolg van endogeniteit is dat het onmogelijk wordt om uitspraken te doen over de causaliteit, in het bijzonder over de grootte van de geschatte effecten.

Mason & Kuhlthau (1992) stelden op basis van *multinomial logit-analyse* van een retrospectieve survey bij 1383 moeders van kinderen op kleuterleeftijd, uitgevoerd in 1986, dat ongeveer 8% van deze vrouwen minder kinderen heeft gekregen dan ze initieel gepland hadden omwille van problemen met de beschikbaarheid, prijs en kwaliteit van plaatselijke kinderopvang. Hun conclusie dat beleid dat gericht is op het aanbieden van voldoende en betaalbare kinderopvang slechts in beperkte mate een invloed heeft op vruchtbaarheidsbeslissingen is problematisch om verschillende redenen. Een eerste probleem is dat enkel vrouwen geselecteerd werden die minstens één schoolgaand kind hebben (kleuterschool). Hierdoor kunnen geen uitspraken gedaan worden over de correlatie tussen kenmerken van kinderopvang en gezinsvorming, aangezien geen vrouwen ondervraagd zijn die besloten hebben om geen kinderen te nemen, rekening houdend met de eventuele beperkte kinderopvang. Een andere bedenking is de vraagstelling: “Hebben problemen met kinderopvang er ooit toe geleid dat u beslist heeft om minder kinderen te krijgen dan u oorspronkelijk van plan was?”[[2]](#footnote-2). De respondent werd hier gevraagd om retrospectief drie verschillende zaken te combineren (intentie, gedrag, kenmerken van kinderopvang), wat meetfouten door *recall bias* niet uitsluit (of zelfs waarschijnlijk maakt). Tot slot merken Rindfuss & Brewster (1996) op dat een stijging van 8% in de context van lage vruchtbaarheid niet gering is.

Hank en Kreyenfeldt (2003) vonden in een *discrete-time hazard* analyse van paneldata (SOEP) van 1984 tot 1999 geen significante verbanden tussen de regionale beschikbaarheid van formele kinderopvang in West-Duitsland en geboorterisico’s voor een eerste of tweede kind voor die periode. De analyses gebeurden onder de assumptie van exogeniteit voor regionale beschikbaarheid van kinderopvang, hetgeen twijfelachtig is. Eerder onderzoek toonde al aan dat de beschikbaarheid van lokale kinderopvang correleert met niet-geobserveerde heterogene kenmerken (Rindfuss et al., 2007). Daarnaast werd beschikbaarheid van opvang voor kinderen van 3 tot 6 jaar gebruikt als proxy voor opvang van jongere kinderen. Deze variabele werd bovendien enkel waargenomen in 1986 en 1994 en aangenomen constant te zijn in de andere periodes (Rindfuss et al, 2007).

Vergelijkbare bedenkingen kunnen gemaakt worden bij het onderzoek van Andersson et al. (2004). Aan de hand van een *event-history analyse* van Zweedse longitudinale data werden slechts kleine en in het licht van de theorie verrassend negatieve verbanden gevonden tussen regionale beschikbaarheid van kinderopvang in 1997 en 1998 en geboorterisico’s voor een tweede of derde kind bij koppels (getrouwd of samenwonend). Ook het positieve verband van die geboorterisico’s met de prijs van kinderopvang en een negatief verband met de kwaliteit ervan gaan in tegen theoretische voorspellingen. Deze bevindingen zijn twijfelachtig aangezien er slechts data beschikbaar was over de regionale beschikbaarheid van kinderopvang gedurende twee jaar. Ook hier is de assumptie van een exogeen aanbod van kinderopvang problematisch (Rindfuss et al., 2007). Daarnaast wordt kinderopvang zeer breed gedefinieerd als opvang voor kinderen met een leeftijd van 1 tot 12 jaar in plaats van specifieke opvang voor jonge kinderen (bijvoorbeeld van 0 tot 3 jaar)[[3]](#footnote-3).

Krapf (2009) rapporteert een eveneens opvallend negatief verband tussen het aantal kinderen (1 of 2 jaar) in kinderopvang in Zweden en gezinsvorming (geboorte van een eerste kind) voor de periode 2001 tot en met 2003. Het gebruik van een logistische regressie met *random effects* laat echter niet toe om te controleren voor mogelijke endogeniteit van de opvangvariabele. Het argument dat endogeniteit in deze studie uitgesloten is aangezien een eerste geboorte geen invloed kan hebben op het regionale gebruik van kinderopvang is eerder zwak. Omgekeerde causaliteit is slechts één van de mogelijke oorzaken van endogeniteit.

Van Bavel & Różańska-Putek (2010) tonen aan de hand van een *discrete-time* *hazard model* een positief, maar zwak verband tussen de interactie van opleidingsniveau van vrouwen en het gebruik van kinderopvang (gemeten als percentage kinderen onder de 3 jaar dat ingeschreven is in formele kinderopvang in 2004) en het geboorterisico voor tweede geboortes. De studie van 16 Europese landen is gebaseerd op retrospectieve gegevens uit cross-sectionele data van de derde ronde van de *European Social Survey* (2008). Zonder controle voor endogeniteit van de beschikbaarheid van kinderopvang en gezien de ruwe gegevens op nationaal niveau van deze variabele is het echter onmogelijk om hieruit causale verbanden te concluderen.

Varga (2014) tot slot, gebruikte een survivalanalyse met *split population design* op basis van retrospectieve surveydata uit Portugal tussen 1960 en 1997. Hij vond enkel een marginaal significant positief verband tussen lokale beschikbaarheid van kinderopvang en kans op een tweede geboorte. Zonder controle voor individuele of regionale heterogeniteit kunnen echter ook hier conclusies uit getrokken worden over de causaliteit. Bovendien wordt kinderopvang gemeten als de proportie kinderen tussen 3 en 5 jaar die ingeschreven zijn in een kleuterschool, wat een betwistbare benadering is voor kinderopvang voor kinderen jonger dan 3 jaar.

Slechts twee onderzoeken houden in hun analyse expliciet rekening met endogeniteit. Baizán (2009) gebruikt als indicator voor kinderopvang de regionale beschikbaarheid (gemeten als het aandeel 0 tot en met 2-jarigen ingeschreven in formele kinderopvang in een bepaalde regio) in een onderzoek naar de keuze voor gezinsvorming (eerste geboorte) en gezinsuitbreiding (tweede en hogere orde geboorten) bij Spaanse vrouwen tussen de 16 en 42 jaar oud. Longitudinale data tussen 1994 en 2001 werd geanalyseerd aan de hand van een *event history* model met regionale *fixed effects*. Onder de assumptie dat regionale kenmerken die zowel de beschikbaarheid van kinderopvang als vruchtbaarheidsbeslissingen beïnvloeden niet veranderen gedurende de geobserveerde periode biedt deze aanpak een oplossing voor *omitted variable bias* in paneldata. Na controle voor verschillende individuele kenmerken voor arbeidsmarkpositie, opleidingsniveau en gezondheid van de vrouw wordt geconcludeerd dat, afhankelijk van de modelspecificatie, het marginale effect van een stijging in beschikbaarheid van kinderopvang zorgt voor een stijging van ongeveer 4,5% tot 7% in de kans op een eerste geboorte en 4,5% tot 6% voor een hogere orde geboorte. Voorts werden geen significante interactie-effecten gevonden voor de opvangvariabele.

Rindfuss et al. (2010) gebruiken een *discrete-time hazard* model voor de schatting van de timing van eerste tot en met vijfde geboorten bij Noorse vrouwen geboren tussen 1957 en 1962. De analyse gebeurde op basis van longitudinale registerdata op individueel en regionaal niveau met zowel vaste variabelen, bijvoorbeeld opleidingsniveau van de ouders, als variabelen die variëren in de tijd, bijvoorbeeld beschikbaarheid van kinderopvang. Deze laatste werd gemeten als het percentage kinderen jonger dan de schoolplichtige leeftijd[[4]](#footnote-4) dat gebruik maakt van kinderopvang, jaarlijks gemeten per regio. Endogeniteit van regionale kinderopvang werd gecontroleerd door gebruik te maken van *fixed effects* op regionaal niveau. Verder werd gecontroleerd voor niet-geobserveerde heterogeniteit door gebruik te maken van een variant op de Heckman-Singer (1984) procedure, waarbij een random term wordt toegevoegd aan het model in de hoop dat dit het effect van weggelaten variabelen vangt (Trussell & Richards, 1985). Op basis van de geschatte parameter werd tot slot een simulatie gemaakt van het gemiddeld aantal kinderen voor vrouwen (tot de leeftijd van 35 jaar) uit geboortecohorten van 1957 tot 1962 en dit voor verschillende niveaus van beschikbaarheid van kinderopvang. De resultaten van deze simulatie tonen een stijging van 0,67 kind bij een toename van de beschikbaarheid van 0 naar 60%.

*Tabel 1* geeft een kort overzicht van de hierboven beschreven empirische literatuur. Wat opvalt is dat, met uitzondering van Mason & Kuhlthau (1992), alle onderzoeken als indicator voor kinderopvang één of andere variant gebruiken van een regionale proportie kinderen dat ingeschreven is in een formele vorm van kinderopvang. Zoals Rindfuss et al. (2010) beargumenteren geeft deze variabele eerder een indicatie van gebruik van kinderopvang dan van beschikbaarheid.

Daarnaast valt op dat de onderzoeken die geen rekening houden met mogelijke endogeniteit uiteenlopende resultaten laten optekenen van het verband tussen kinderopvang en vruchtbaarheid. Beide onderzoeken die controleren voor endogeniteit tonen daarentegen een sterk positief effect, wat consistent is met de theoretische voorspellingen. Uiteraard is het voorbarig om hieruit te concluderen dat dit louter het resultaat is van deze controle, geen van beiden heeft bijvoorbeeld nagegaan wat de verandering is in de geschatte parameters tussen een modelspecificatie met of zonder controle voor endogeniteit.

Tot slot maken beide causale onderzoeken (Baizàn, 2009; Rindfuss et al., 2010) gebruik van een fixed-effectsmodel om te controleren voor endogeniteit. Deze techniek heeft enkele belangrijke nadelen. Ten eerste laat ze niet eenvoudigweg toe om variabelen op te nemen die niet variëren in de tijd, hetgeen problematisch kan zijn voor onderzoekers die enkel beschikken over cross-sectionele data voor bepaalde kenmerken. Ten tweede staat of valt de validiteit ervan met de assumptie dat niet geobserveerde kenmerken onveranderlijk zijn gedurende de geobserveerde periode. Het is echter niet ondenkbaar dat bepaalde niet-geobserveerde kenmerken systematisch kunnen variëren met regionale kenmerken.

## 3. Kinderopvang in Vlaanderen

In dit onderzoek wordt elke vorm van opvang die niet gebeurt door één van de ouders beschouwd als kinderopvang. Op huishoudensniveau wordt de zorg door een ander gezinslid, zoals een inwonende grootouder of een ouder kind dus beschouwd als kinderopvang. Om het onderscheid te maken met opvang buiten het gezin wordt dit aangeduid met *interne opvang*.

Wanneer ouders ervoor opteren gebruik te maken van externe opvang, kan dit op verschillende manieren. Opvang kan *informeel* gebeuren door niet-inwonende grootouders of een ander familielid of kan overgelaten worden aan derden. De opvang kan thuis (au pair, babysit, huispersoneel,… ) of buitenshuis gebeuren. Voor kinderen die buitenshuis *formeel* door derden opgevangen worden bestaat er het onderscheid tussen gezinsopvang of groepsopvang.

Onthaalouders (gezinsopvang) zijn ofwel aangesloten zijn bij een erkende dienst voor opvanggezinnen, of werken op zelfstandige basis. Voor de eerste groep zijn de tarieven wettelijk bepaald en inkomensafhankelijk. Groepsopvang kan opgedeeld worden in erkende en zelfstandige kinderdagverblijven. Voor de eerste groep geldt dat er wettelijke kwaliteitsnormen (opleiding van begeleiders, …) en inkomensafhankelijke prijzen zijn.[[5]](#footnote-5)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Auteurs | Type data | Methode | Indicator vruchtbaarheid | Indicator kinderopvang | Resultaat |
| Mason & Kuhlthau (1992) | Cross-sectionele Survey (1986) | Multinomial logit | Zelfgerapporteerd | Zelfgerapporteerd, beperkingen in kinderopvang | Negatief verband |
| Hank & Kreyenfeldt (2003) | Longitudinaal (SOEP)  (1984 - 1999) | Discrete-time hazard | Pariteit: Eerste / tweede kind | Regionale beschikbaarheid | Geen significante verbanden |
| Andersson et al. (2004) | Longitudinaal  (1997 – 1998) | Event history | Pariteit: Tweede / derde | Regionale beschikbaarheid | Negatieve verbanden |
| Krapf (2009) | Longitudinaal  (2001 – 2003) | Logistische regressie, random effects | Pariteit: Eerste | Regionale beschikbaarheid | Negatief verband |
| Van Bavel & Różańska-Putek (2010) | Longitudinaal Retrospectief  (ESS 2008) | Discrete-time hazard | Pariteit: Tweede | Regionale beschikbaarheid (interactie met opleiding vrouw) | Positief verband |
| Varga (2014) | Longitudinaal  Retrospectief  (1960 – 1997) | Split population | Pariteit: Eerste, tweede, derde en hoger | Regionale beschikbaarheid kinderopvang | Positief (enkel voor tweede geboorte) |
| Baizán (2009) | Longitudinaal  (1994 – 2001) | Event history  fixed effects | Pariteit: Eerste / tweede en hoger | Regionale beschikbaarheid kinderopvang | Positief effect |
| Rindfuss et al. (2010) | Longitudinaal  (1993 – 1978) | Event history  Fixed effects | Pariteit: Eerste t.e.m. vijfde | Regionale beschikbaarheid kinderopvang | Positief effect |

Tabel 1: Overzicht empirische literatuur

De belangrijkste vormen van gebruikte kinderopvang in België zijn die door grootouders, erkende onthaalouders en de erkende kinderdagverblijven (Hedebouw, 2009). Opvallend is dat het aandeel in de opvang door grootouders met een derde verminderd is tussen 2002 en 2009 (van 34,4% naar 22,4%), terwijl de twee andere vormen allebei gestegen zijn, respectievelijk van 26,3% naar 28,5% en van 15,6% naar 18,4%. Uit de cijfers van Hedebouw blijkt voorts dat de zelfstandige onthaalouders en kinderdagverblijven samen in 2002 een kleine 20% van de opvang voor hun rekening namen.

Naast de verschillende mogelijkheden voor kinderopvang moet er ook rekening gehouden worden met heterogeniteit qua intensiteit van die opvang. Voor 2002 vond Hedebouw (2009) dat 52,2% van de ouders van jonge kinderen (3 maand tot 3 jaar) in het Vlaams Gewest regelmatig gebruik maakt van kinderopvang, tegenover 11,1% met een beperkt gebruik[[6]](#footnote-6). In 2009 was het regelmatig gebruik gestegen tot 63,2%, terwijl het beperkt gebruik terugviel tot 5,8%.

De keuze voor de verschillende vormen van kinderopvang is daarnaast niet disjunct. Een beperkt gebruik van formele kinderopvang wil bijvoorbeeld niet zeggen dat er een beperkt gebruik is van kinderopvang an sich. Bovendien geldt dat de verschillende vormen van kinderopvang zowel complementen als substituten kunnen zijn. Wanneer iemand de overgang maakt van deeltijds naar voltijds werken, is het niet ondenkbaar dat zowel het gebruik van formele als informele kinderopvang zal stijgen. Wanneer daarnaast de prijs van formele kinderopvang stijgt, kan het zijn dat deze (deels) vervangen wordt door informele opvang.

Ouders kunnen er tot slot ook voor kiezen om geen enkele vorm van niet-ouderlijke opvang te voorzien en er volledig zelf voor in te staan. In de eerste plaats wordt hierbij gedacht aan één van de partners die de arbeidsmarkt verlaat (of kiest om ze niet te betreden) om voor de kinderen te zorgen. Voor Vlaanderen blijkt dat ongeveer een derde van de ouders van kinderen met een leeftijd tussen 3 maand en 3 jaar ervoor kiest om geen gebruik te maken van kinderopvang, al is hun aandeel wel gedaald van 36,7% in 2002 tot 31% in 2009 (Hedebouw, 2009).

## 4. Hypothesen

Op basis van het hiervoor beschreven overzicht van de theoretische en empirische literatuur, en de situatieschets van kinderopvang in Vlaanderen, worden in het empirische gedeelte van deze thesis de onderzoeksvraag: *wat is het effect van kinderopvang op vruchtbaarheid?,* onderzocht aan de hand van volgende hypothesen:

Aangenomen dat de prijs van kinderopvang in de meeste gevallen lager is dan het loon dat vrouwen kunnen verdienen bij participatie op de arbeidsmarkt (Rindfuss, 2010) en het gebruik van kinderopvang daarom een manier is om de opportuniteitskosten van kinderen te verlagen en tegelijkertijd het gezinsinkomen te verhogen (Becker, 1981) wordt volgende hypothese getoetst:

1. Het gebruik van formele kinderopvang heeft een positief effect op de kans op een tweede geboorte.

Gelet op het feit dat de opportuniteitskosten van kinderen groter zijn voor hoger opgeleide vrouwen en vrouwen met betere arbeidsmarktvooruitzichten tout court, luidt de tweede hypothese:

1. Het effect van het gebruik van formele kinderopvang op de kans op een tweede geboorte is groter voor
   1. Hoger opgeleide vrouwen
   2. Vrouwen met betere arbeidsmarktkansen

Aangezien het gebruik van informele kinderopvang gezien kan worden als een (goedkoper) substituut voor formele kinderopvang, wordt als derde hypothese gesteld:

1. Het effect van formele kinderopvang op de kans op een tweede geboorte is kleiner voor vrouwen die gebruik maken van informele opvang.

Gezien de theorie rond endogene voorkeuren en de rol van sociale context (Lesthaeghe & Surkyn, 1988) in vruchtbaarheidsbeslissingen, waarbij gedrag beïnvloed wordt door het gedrag en attitudes van anderen (Kohler, 2001 volgens Bongaarts, 2002), luidt de vierde en laatste hypothese:

1. Het effect van het gebruik van kinderopvang is robuust voor regionale kenmerken die verband houden met de kans op een tweede geboorte.

# Deel 2: Empirische analyse

## 5. Methodologie & Data

### 5.1 Methodologie

#### 5.1.1 endogeniteit, problemen en oplossingen

Zoals reeds bleek uit de studie van literatuur en eerder empirisch onderzoek, is het waarschijnlijk dat onderzoek naar de invloed van het gebruik van kinderopvang op de keuze voor een tweede kind bemoeilijkt wordt door mogelijke endogeniteitsproblemen. Eerder werd al vermeld dat Hill et al. (2012) drie belangrijke oorzaken van endogeniteit beschrijven.

Een eerste mogelijke oorzaak is wanneer de onafhankelijke variabelen meetfouten bevatten. Voor dit specifieke onderzoek is deze vorm van endogeniteit echter onwaarschijnlijk, zoals later uit de bespreking van de data zal blijken.

Als tweede mogelijke oorzaak is er omgekeerde causaliteit of *simultaneïteitssbias*. Dit probleem stelt zich wanneer de onafhankelijke en afhankelijke variabele elkaar wederzijds beïnvloeden. Eerder beschreven onderzoek gebruikt voornamelijk (met uitzondering van Mason & Kuhlthau, 1992) regionale beschikbaarheid van kinderopvang als indicator. Deze wordt gemeten door de proportie kinderen van een bepaalde leeftijdscategorie die gebruik maken van een vorm van kinderopvang. Wanneer er echter in een bepaalde regio meer kinderen geboren worden, kan het zijn dat de beschikbaarheid verandert. Een andere mogelijkheid is dat private kinderopvang anticipeert op een verwachte stijging van het aantal geboorten en zo meer plaatsen creëert.

De belangrijkste oorzaak waarmee rekening gehouden moet worden is *omitted variable bias*, in wezen een catch-all term die erop duidt dat bepaalde kenmerken die in verband staan met zowel kinderopvang als vruchtbaarheid niet geobserveerd werden, wat opgelost kan worden door uitgebreidere data, of (bijna) niet observeerbaar zijn, zoals persoonlijke voorkeur. Vaak wordt deze oorzaak van endogeniteit aangeduid met *unobserved heterogeneity*.

Niet geobserveerde heterogeniteit kan voorkomen op elk meetniveau. Wanneer gekeken wordt naar de individuele keuze om kinderen te krijgen en al dan niet gebruik te maken van kinderopvang is het aannemelijk dat niet geobserveerde factoren beide variabelen kunnen beïnvloeden. Hetzelfde geldt voor indicatoren op bijvoorbeeld regionaal niveau. Indien er bepaalde regionale kenmerken zijn die een invloed hebben op vruchtbaarheidsbeslissingen, hebben deze waarschijnlijk eveneens een invloed op de regionale beschikbaarheid van kinderopvang (Rindfuss et al., 2007).

De belangrijkste gevolgen van endogeniteit zijn dat statistische schattingen niet langer consistent zijn en bijgevolg geen uitspraak kan gedaan worden over de precieze grootte van een verband. Bovendien wordt het onmogelijk om precieze causaliteit te bepalen. Aangezien endogeniteit een mogelijk probleem is bij elk niet-experimenteel onderzoek bieden specifieke analysetechnieken een uitkomst.

*Fixed effects*-modellen, zoals gebruikt in de onderzoeken van Baizàn (2009) en Rindfuss et al. (2010), bieden een uitkomst voor longitudinale of multilevel-analyses. De belangrijkste nadelen zijn dat ze niet eenvoudig toelaten om variabelen te analyseren die niet veranderen doorheen de tijd en dat er mogelijk problemen zijn met de assumptie dat de niet-geobserveerde kenmerken constant zijn overheen de verschillende observatiemomenten. Zoals verderop uit de databespreking zal blijken, maakt dit onderzoek voornamelijk gebruik van cross-sectionele data op individueel niveau, waardoor deze aanpak niet toegepast kan worden.

Er werd daarom geopteerd om gebruik te maken van *instrumentele variabelen* (IV). De kracht van deze econometrische aanpak is dat ze een oplossing beidt voor alle drie de beschreven oorzaken van endogeniteit (Angrist & Pischke, 2009). In essentie wordt in een IV-model gebruik gemaakt van variabelen die gecorreleerd zijn met de afhankelijke variabele, maar enkel omdat ze gecorreleerd zijn met de onafhankelijke variabele van belang. Het meest voor de hand liggende bezwaar tegen deze techniek is dat het in de praktijk een uitdaging blijkt om geschikte instrumenten te vinden (Aassve & Lappegård; 2009).

Een tweede voordeel van deze techniek is dat veel statistische softwarepakketten ingebouwde methoden hebben om gebruik te maken van instrumentele variabelen, afhankelijk van het onderzoeksopzet. Een belangrijke beperking is dat zo’n schattingsmethoden niet bestaan voor de logistische modellen die, omwille van de voordelen op het vlak van interpretatie, alomtegenwoordig zijn in event-history analyse. Er werd daarom gekozen voor een probit-model, waarvoor wel schattingstechnieken voorhanden zijn.

Aangezien het gebruik van niet-traditionele methoden ongetwijfeld een onderwerp is van discussie (overigens terecht), wordt in de appendix (A) op een pragmatische manier geargumenteerd waarom deze aanpak gerechtvaardigd is.

#### 5.1.2 het probit-model met instrumentele variabelen

De keuze voor een logistische regressie is in zekere mate arbitrair (Allison, 1982), waarbij het grootste voordeel ligt in de relatief eenvoudige interpretatie van de coëfficiënten. Een ander model waarbij de afhankelijke variabele binair is, is het *probit-model*.

De kans op een geboorte in periode *t* wordt hierbij voorgesteld als volgt:

 (1)

Met de afhankelijke variabele die waarde 1 aanneemt bij een geboorte op het moment *t*, en 0 wanneer geen geboorte plaatsvindt. De kans op een geboorte (*Pit*) wordt dan voorgesteld door een functie *F(zit)*, waarbij *zit* op zijn beurt een functie is van tijdsindicatoren , onafhankelijke variabelen  en een vector parameters .

Het econometrische model wordt dan afgeleid van een latente variabele  die het netto-nut van het krijgen van een tweede kind voorstelt en aangenomen wordt lineair te zijn:

 (2)

We observeren , de keuze voor een tweede kind, wanneer , het netto nut, positief is.

 (3)

Wanneer we aannemen dat de foutenterm () standaardnormaal verdeeld is, kan het probitmodel als volgt voorgesteld worden, met de cumulatieve kansdichtheidsfunctie:

 (4)

Te herleiden tot:  (5)

Wanneer één van de onafhankelijke variabelen endogeen is, herschrijven we vergelijking (2) als volgt:

 (6)

Waarbij de parameter voorstelt die hoort bij , de endogene variabele die zelf voorgesteld wordt als een functie van exogene variabelen *zi* en hun bijhorende parameters :

 (7)

In eerste stap worden geschatte waardenberekend op basis van (7), die in een tweede stap de waarden van de oorspronkelijke variabele vervangen zodat (6) herschreven kan worden in structurele vorm:

 (8)

### 5.2 Data

Voor de analyse van het effect van formele kinderopvang werd een selectie gemaakt van Vlaamse vrouwen met één eigen jong kind (0 tot 2 jaar) uit de Belgische census van 2001. Deze vrouwen werden op basis van rijksregistergegevens jaarlijks prospectief opgevolgd tot maximaal 2005 of het moment waarop een tweede geboorte plaatsvond. Voorts werd de dataset gekoppeld aan data van Kind en Gezin die gegevens bevat over regionale karakteristieken van kinderopvang in Vlaanderen[[7]](#footnote-7).

Als afhankelijke variabele wordt gekeken naar de aanwezigheid van een tweede geboorte in de jaren volgend op de census tot maximaal 2005. Vrouwen waarvoor problematische gegevens werden vastgesteld, werden verwijderd. Voorbeelden hiervan zijn vrouwen waarvoor de uitkomst van de berekening “leeftijd – leeftijd bij eerste kind” een negatief getal opleverde. Voorts zijn ook vrouwen die in de census aangaven meer dan 1 kind te hebben in kinderopvang, kleuterschool,… verwijderd. Een mogelijke verklaring is dat het hier om niet-biologische kinderen gaat van vrouwen met één eigen kind, bijvoorbeeld in nieuw-samengestelde gezinnen of bij adoptie. Om te vermijden dat het gebruik van kinderopvang vertekend wordt in de periode vlak voor of na een geboorte, werden ook alle vrouwen met een eerste kind geboren in het jaar van de census (2001) geweerd. Deze selecties resulteren uiteindelijk in 32.434 vrouwen die maximaal 6 jaar gevolgd werden (geboorte eerste kind in 1999), wat uiteindelijk resulteert in 88.379 persoonsjaren gedurende welke deze vrouwen worden geobserveerd en waarbij 18.801 tweede geboorten werden vastgesteld.

De voornaamste onafhankelijke variabele is het gebruik van formele kinderopvang voor het eerste kind, geregistreerd op het moment van de census (2001). In het model met instrumentele variabelen worden drie instrumenten gebruikt:

1. **De gemeentelijke proportie van gesubsidieerde kinderopvang:** De redenering hierbij is dat kinderopvang gemiddeld genomen goedkoper is in gemeenten waar een groter aanbod is van relatief goedkopere gesubsidieerde kinderopvang dan in gemeenten waar er een grotere proportie niet-gesubsidieerde kinderopvang is. Deze variabele heeft geen rechtstreeks effect op de keuze voor een tweede kind, maar bepaalt wel de keuze voor het gebruik van formele kinderopvang door het prijsmechanisme.
2. **Tevredenheid over plaatselijke kinderopvang:** Hierbij wordt aangenomen dat subjectieve onvrede met plaatselijke kinderopvang een negatief effect heeft op het gebruik ervan, maar niet rechtstreeks gecorreleerd is met de keuze voor een tweede kind. De variabele is gecodeerd als dummyvariabele waarbij 1 staat voor ontevredenheid met de plaatselijke opvangvoorzieningen.
3. **Aantal inwonende volwassenen**: Hiermee wordt bedoeld het aantal personen ouder dan 18 jaar die aanwezig zijn in het huishouden. Deze variabele is een indicatie voor eventuele substituten voor formele kinderopvang. Het gaat om een dummyvariabele die de waarde 1 aanneemt wanneer er meer dan 2 personen ouder dan 18 jaar in het gezin wonen.

Als controlevariabelen werden in een eerste fase individuele kenmerken toegevoegd die een effect kunnen hebben op enerzijds het gebruik van formele kinderopvang en anderzijds de keuze voor een tweede geboorte. In de eerste plaats het gebruik van informele opvang, omdat dit een (goedkoper) substituut is voor formele opvang. Daarnaast zijn er leeftijd bij het eerste kind (en het kwadraat), opleidingsniveau (lager – midden – hoger), gezondheidstoestand (om te controleren voor langdurige ziekte) en herkomst (een dummyvariabele voor autochtoon of niet). Al deze variabelen zijn gemeten op het moment van de census (2001).

Aangezien tewerkstelling theoretisch sterk samenhangt met zowel het gebruik van formele kinderopvang als de kans op een tweede geboorte, maar de actuele tewerkstellingsstatus in jaar *x* niet noodzakelijk dezelfde is als die op het moment van de census, werd ervoor geopteerd om tewerkstellingskansen te schatten (op basis van censusgegevens). Deze schattingen gebeurden aan de hand van een probitmodel met als gebruikte variabelen leeftijd (lineair en kwadratisch), opleidingsniveau, de herkomstvariabele en de regionale proportie deeltijds werk. Schattingen van de tewerkstellingsstatus in 2001 op basis van deze regressie resulteerde in 88,73% correcte classificering.

Variabelen die initieel in aanmerking kwamen, maar waarvoor een grote kans bestond dat ze tussen 2001 en (maximaal) 2005 veranderd zouden zijn, werden niet weerhouden. Het gaat hier om gezinskenmerken zoals burgerlijke staat en werkstatus van de partner op het moment van de census.

Naast individuele kenmerken spelen ook omgevingskenmerken een rol in de vruchtbaarheidsbeslissingen. Om de hypothese te toetsen dat gedrag beïnvloed wordt door het gedrag en attitudes van anderen (Kohler, 2001 volgens Bongaarts, 2002), werden twee regionale variabelen opgenomen. Beiden meten in zekere zin hoe “jong” de gemeente van de respondent is, maar op verschillende momenten. De eerste variabele is de proportie geboorten van een bepaalde gemeente in 2001 (aantal geboorten / aantal inwoners). Deze variabele is een indicatie van de vruchtbaarheidsbeslissingen van vrouwen uit dezelfde gemeente. De tweede regionale variabele is de gemeentelijke proportie kleuters (aantal kleuters / aantal inwoners). Gemeenten waar meer kleuters wonen worden geacht “kindvriendelijker” te zijn dan anderen. Wanneer in een gemeente bovendien veel jonge kinderen wonen, zouden vrouwen volgens de theorie van sociale interactie meer geneigd zijn voor kinderen te kiezen.

De proportie deeltijds werk is het aandeel van de tewerkgestelden in een bepaalde gemeente dat in 2001 minder dan 32 uren per week werkte. Deze variabele geeft een indicatie voor de arbeidsmarktkansen van vrouwen, die gemiddeld vaker deeltijds werken dan mannen. *Tabel 2* geeft een beschrijving van de gebruikte variabelen.

Tabel : Beschrijving van de gebruikte variabelen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variabele** | | **% binnen groep**  **(N=32.434)** | | | |
| **Geboorte** | *ja* | 57,97% | | | |
| *nee* | 42,03% | | | |
| **Gebruik formele**  **kinderopvang** | *ja* | 59,48% | | | |
| *nee* | 40,52% | | | |
| **Gebruik informele**  **kinderopvang** | *ja* | 40,32% | | | |
| *nee* | 59,68% | | | |
| **Opleidingsniveau** | *Lager* | 10,71% | | | |
| *Midden*  *Hoger* | 38,27%  51,02% | | | |
| **Slechte gezondheid** | *Ja* | 0,72% | | | |
| *nee* | 99,28% | | | |
| **autochtoon** | *ja* | 87,85% | | | |
| *nee* | 12,15% | | | |
| **Geboorte eerste kind** | *1999* | 37,27% | | | |
| *2000* | 62,73% | | | |
| **Meer dan 2 volwassenenb** | *ja* | 1,50% | | | |
| *nee* | 98,50% | | | |
| **Ontevredenheid plaatselijke**  **Kinderopvangb** | *ja* | 19,53% | | | |
| *nee* | 80,47% | | | |
|  | | **Gemiddelde** | **St.afw** | **Min.** | **Max.** |
| **Leeftijd (lineair)a (N=88.379)** | | 30,87 | 4,49 | 18 | 52 |
| **Leeftijd (kwadratisch)a (N=88.379)** | | 972.83 | 286,54 | 324 | 2704 |
| **Leeftijd bij eerste kind (lineair)** | | 28,01 | 4,01 | 16 | 46 |
| **Leeftijd bij eerste kind (kwadratisch)** | | 800,6 | 231,62 | 256 | 2116 |
| **Tewerkstellingskans (N=88.379)** | | 0,8769 | 0,12 | 0,0182 | 0,9761 |
| **Gemeentelijke proportie deeltijds werk (2001)** | | 0,195 | 0,017 | 0,1547 | 0,255 |
| **Gemeentelijke proportie geboorten (2001)** | | 0,006 | 0,0009 | 0,001 | 0,01 |
| **Gemeentelijke proportie kleuters (2001)** | | 0,1089 | 0,012 | 0,0713 | 0,1519 |
| **Proportie gesubsidieerde opvang (2001)b** | | 0,8451 | 0,14 | 0 | 1 |

a: variabele exclusief gebruikt voor de schatting van tewerkstellingskansen  
b: variabele exclusief gebruikt als instrument voor formele kinderopvang

## 6. Resultaten van de analyses

### 6.1 Beschrijvende analyse: het gebruik van kinderopvang

In het volgende deel wordt aan de hand van enkele illustraties het gebruik van kinderopvang door de vrouwen in de dataset geschetst. Alle figuren zijn berekend aan de hand van censusdata, dus de toestand in 2001. *Figuur 1* geeft het gebruik van formele kinderopvang weer uitgesplitst naar opleidingsniveau. Hieruit blijkt dat hoger opgeleiden vaker gebruik maken van formele kinderopvang dan lager- en middelbaar opgeleiden.

Figuur : Gebruik van formele kinderopvang, naar opleidingsniveau

Bron: Census 2001; eigen bewerking

Een verdere analyse van de opleidingsgradiënt leert daarnaast dat hoger opgeleide vrouwen minder vaak gebruik maken van enkel informele opvang, maar vaker van een combinatie van de twee dan vrouwen uit de andere opleidingscategorieën (*Figuur 2)*.

Figuur : Gebruik van zowel formele als informele opvang naar opleidingsniveau

Bron: Census 2001; eigen bewerking

Het mag duidelijk zijn dat opleiding een belangrijke rol speelt in het gebruik van kinderopvang. Volgens de economische theoretische verklaringen is dit omdat hoger opgeleide vrouwen waarschijnlijk een hoger loon kunnen verdienen, waardoor zelf instaan voor de zorg een grotere opportuniteitskost zou betekenen. Daarnaast speelt een inkomenseffect, aangenomen dat homogamie ook van toepassing is op deze vrouwen zal het gezinsinkomen van hoger opgeleide vrouwen waarschijnlijk groter zijn dan dat van vrouwen met een lagere opleiding. Hierdoor is de relatieve kost van kinderopvang kleiner.

Eveneens voorspelbaar is de uitsplitsing van het gebruik van formele kinderopvang naar werkende en werkloze vrouwen (*figuur 3*). Van de werkende vrouwen (in 2001) geeft meer dan 60% aan dat ze gebruik maken van formele kinderopvang. Daarnaast geeft meer dan 15% van de vrouwen aan niet te participeren op de arbeidsmarkt, maar wel gebruik te maken van formele kinderopvang.

Figuur : Gebruik van formele kinderopvang naar arbeidsmarktparticipatie

Bron: Census 2001; eigen bewerking

Uit *figuur 4* blijkt dat het gebruik van formele kinderopvang het grootst is bij vrouwen tussen de 30 en de 40 jaar oud. 60% van de jonge moeders van 19 tot 24 jaar oud geven aan geen gebruik te maken van formele kinderopvang.

Figuur : Gebruik van formele kinderopvang naar leeftijd

Bron: Census 2001; eigen bewerking

Figuur : Gebruik van formele kinderopvang naar herkomst

Bron: Census 2001; eigen bewerking

Er is minder verschil te merken in het gebruik van formele kinderopvang door autochtone en allochtone vrouwen. De eerste groep geeft in iets meer dan 60% van de gevallen aan gebruik te maken, tegenover ongeveer 50% van de allochtone vrouwen (*figuur 5*).

Tot slot van deze beschrijvende analyse blijkt uit *figuur 6* dat bij 55% van de vrouwen die aangeeft gebruik te maken van formele kinderopvang voor een eerste kind aan het einde van de geobserveerde periode een tweede geboorte werd vastgesteld, tegenover 45% bij vrouwen die geen gebruik maken van formele kinderopvang.

Figuur : Tweede geboorten en gebruik van formele kinderopvang

Bron: Census 2001; eigen bewerking

### 6.2 Verklarende analyse: het effect van kinderopvang

*Tabel 3* geeft de resultaten weer van zes verschillende regressies voor drie verschillende modelspecificaties. Elke modelspecificatie werd geschat door een eenvoudige probitregressie en daarnaast aan de hand van een *IVprobit*, een probitregressie die gebruik maakt van instrumentele variabelen. In deze regressies werd de variabele die het gebruik van formele kinderopvang voor het eerste kind weergeeft geïnstrumenteerd door de gemeentelijke proportie gesubsidieerde kinderopvang, het aantal inwonende volwassenen en de perceptie van de lokale kinderopvang. Resultaten van het eerste stadium van de regressies met instrumentele variabelen kan teruggevonden worden in de appendix (C).

De significantie van het model wordt gemeten aan de hand van de *Wald Chi-kwadraat*-toets. Deze toont dat alle zes de modellen significant zijn. Aan de hand van een *LR-test* (Likelihood Ratio test) werd gekeken of de uitbreidingen van de modellen die met instrumentele variabelen geschat werden een significante verbetering zijn ten opzichte van het vorige model. Uit de resultaten blijkt dat dit het geval is.

De *Wald test of exogeneity* is een chi-kwadraattoets die de nulhypothese toetst dat de geïnstrumenteerde variabele exogeen is. Verwerpen van deze hypothese is daarom een indicatie dat de variabele endogeen is en dat het gebruik van instrumentele variabelen aangewezen was. Uit de resultaten blijkt dat bij alle drie de modelspecificaties het gebruik van instrumentele variabelen gerechtvaardigd was, met *p < 0,005* voor het eerste model en *p < 0,05* voor de twee anderen. Resultaten van aanvullende toetsen die de relevantie en validiteit van de gebruikte instrumenten beoordelen, alsook een uitgebreide bespreking ervan, zijn eveneens terug te vinden in de appendix (B).

Resultaten van probitanalyses zijn niet eenvoudig te interpreteren, uitgezonderd het teken. Negatieve parameters wijzen op een negatief verband en omgekeerd. Om een algemeen beeld te krijgen werd daarom ook het gemiddelde marginale effect van formele kinderopvang weergegeven voor elk van de modellen. Dit marginale effect geeft weer wat de procentuele stijging in de kans op een tweede geboorte, bij een eenheidsverandering van het gebruik van formele kinderopvang is per jaar. Om de interpretatie te vereenvoudigen is dit effect geschat bij niet-gebruik van formele kinderopvang, een eenheidsverandering komt daarom neer op het wél gebruik maken van formele kinderopvang.

*Tabel 4* geeft daarnaast een overzicht van geschatte marginale effecten van het gebruik van formele kinderopvang bij verschillende waarden voor de andere onderzochte variabelen. Deze effecten zijn berekend aan de hand van de resultaten van de *IVprobit*-regressie van het derde model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Tabel 3: Resultaten van probitregressies* | | | | | | |
| covariaten | Model 1 (probit) | Model 1 (IVprobit) | Model 2 (probit) | Model 2 (IVprobit) | Model 3 (Probit) | Model 3  (IVprobit) |
| Tijdsindicatoren late entry (ref = 2) |  |  |  |  |  |  |
| 3 | -0,02+  (0,012) | -0,008  (0,013) | 0,009  (0,013) | 0,01  (0,013) | 0,01  (0,013) | 0,01  (0,013) |
| 4 | -0,34\*\*\*  (0,014) | -0,30\*\*\*  (0,017) | -0,28\*\*\*  (0,014) | -0,27\*\*\*  (0,016) | -0,28\*\*\*  (0,014) | -0,26\*\*\*  (0,016) |
| 5 | -0,66\*\*\*  (0,016) | -0,61\*\*\*  (0,022) | -0,59\*\*\*  (0,017) | -0,56\*\*\*  (0,022) | -0,58\*\*\*  (0,017) | -0,56\*\*\*  (0,022) |
| 6 | -0,89\*\*\*  (0,026) | -0,83\*\*\*  (0,031) | -0,81\*\*\*  (0,026) | -0,78\*\*\*  (0,032) | -0,81\*\*\*  (0,026) | -0,78\*\*\*  (0,032) |
| constante | -0,67\*\*\*  (0,011) | -0,94\*\*\*  (0,056) | -2,78\*\*\*  (0,206) | -2,7\*\*\*  (0,21) | -3,15\*\*\*  (0,213) | -3,2\*\*\*  (0,222) |
|  |  |  |  |  |  |  |
| Individuele kenmerken |  |  |  |  |  |  |
| Formele kinderopvang (ref=nee) | 0,16\*\*\*  (0,01) | 0,65\*\*\*  (0,104) | 0,063\*\*\*  (0,012) | 0,63\*\*  (0,212) | 0,063\*\*\*  (0,012) | 0,61\*\*  (0,213) |
| Informele kinderopvang (ref=nee) |  |  | -0,072\*\*\*  (0,012) | 0,22\*  (0,113) | -0,071\*\*\*  (0,012) | 0,21+  (0,114) |
| Leeftijd bij eerste kind |  |  | 0,20\*\*\*  (0,019) | 0,19\*\*\*  (0,021) | 0,18\*\*\*  (0,02) | 0,16\*\*\*  (0,022) |
| Leeftijd bij eerste kind (kwadraat) |  |  | -0,005\*\*\*  (0,0003) | -0,004\*\*\*  (0,0004) | -0,004\*\*\*  (0,0004) | -0,0036\*\*\*  (0,0004) |
| Opleidingsniveau (ref=lager) |  |  |  |  |  |  |
| Hoger |  |  | 0,61\*\*\*  (0,033) | 0,50\*\*\*  (0,058) | 0,55\*\*\*  (0,035) | 0,45\*\*\*  (0,058) |
| Midden |  |  | 0,15\*\*\*  (0,024) | 0,12\*\*\*  (0,027) | 0,11\*\*\*  (0,025) | 0,08\*\*\*  (0,028) |
|  |  |  |  |  |  |  |
| Tewerkstellingskans |  |  | -0,43\*\*\*  (0,165) | -0,55\*\*  (0,167) | -0,09  (0,175) | -0,2  (0,177) |
| Slechte gezondheid (ref=nee) |  |  | -0,32\*\*\*  (0,064) | -0,18\*  (0,082) | -0,31\*\*\*  (0,064) | -0,18\*  (0,082) |
| Autochtoon (ref=nee) |  |  | 0,06\*\*\*  (0,037) | 0,027  (0,039) | 0,0016  (0,038) | -0,032  (0,04) |
|  |  |  |  |  |  |  |

Tabel 3 (vervolg)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| covariaten | Model 1 (probit) | Model 1 (IVprobit) | | Model 2 (probit) | Model 2  (IVprobit) | Model 3  (Probit) | | Model 3  (IVprobit) |
| Regionale kenmerken |  |  | |  |  |  | |  |
| Proportie geboorten |  |  | |  |  | 4,25  (6,679) | | 0,04  (6,807) |
| Proportie kleuters |  |  | |  |  | 3,53\*\*\*  (0,478) | | 3,40\*\*\*  (0,479) |
| Proportie deeltijds werk |  |  | |  |  | 1,71\*\*\*  (0,318) | | 1,74\*\*\*  (0,314) |
|  |  |  | |  |  |  | |  |
| Gemiddeld Marginaal effect van formele kinderopvang | 0,0410\*\*\*  (0,002) | 0,1309\*\*\*  (0,014) | | 0,0165\*\*\*  (0,003) | 0,1250\*\*\*  (0,029) | 0,0163\*\*\*  (0,003) | | 0,1220\*\*\*  (0,03) |
| Modeltoetsen |  |  | |  |  |  | |  |
| Log likelihood |  | -106663,75 | |  | -87891,422 |  | | -87822,035 |
| Wald Chi² (df) |  | 3693,84 (5) \*\*\* | |  | 6851,84 (13)\*\*\* |  | | 6884,44 (16)\*\*\* |
|  |  |  | |  |  |  | |  |
| LR-test Chi² (df) |  | - | |  | 377544,66 (16)\*\*\* |  | | 1238,77 (6)\*\*\* |
|  |  |  | |  |  |  | |  |
| Wald test of exogeneity Chi² (df) |  | 20,23 (1)\*\*\* | |  | 6,51(1)\* |  | | 6,04(1)\* |
|  |  | |  |  |  | |  |  |

Standaardfouten tussen haakjes; + p < 0,1 ; \* p<0,05 ; \*\* p<0,01 ; \*\*\*p<0,005; gemiddelde marginale effecten van formele kinderopvang zijn berekend bij formeel=0

Tabel : Marginale effecten van het gebruik van formele kinderopvang bij gebruik = 0

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| variabele | waarde | dy/dx | S.E. | 95% C.I. | |
| Gemiddeld marginaal effect  van formele kinderopvang |  | 0,12 | 0,03 | 0,06 | 0,18 |
| Informele opvang | **Nee** | 0,11 | 0,02 | 0,07 | 0,15 |
|  | **Ja** | 0,13 | 0,04 | 0,05 | 0,22 |
| Leeftijd bij eerste kind |  |  |  |  |  |
| Min | **16** | 0,01\* | 0,004 | 0,001 | 0,02 |
| Q1 | **25** | 0,09 | 0,02 | 0,04 | 0,13 |
| Med | **28** | 0,13 | 0,04 | 0,06 | 0,2 |
| Q3 | **31** | 0,16 | 0,05 | 0,06 | 0,26 |
| Max | **46** | ***0,08*** | 0,05 | -0,02 | 0,19 |
| Opleidingsniveau | **Hoger** | 0,15 | 0,04 | 0,08 | 0,22 |
|  | **Midden** | 0,10 | 0,03 | 0,05 | 0,15 |
|  | **Lager** | 0,09 | 0,02 | 0,05 | 0,14 |
| Tewerkstellingskans |  |  |  |  |  |
| Min | **1,82%** | 0,14 | 0,05 | 0,05 | 0,23 |
| Q1 | **85,06%** | 0,13 | 0,03 | 0,06 | 0,18 |
| Med | **91,26%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Q3 | **96,26%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Max | **97,61%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Slechte Gezondheid | **Nee** | 0,12 | 0,03 | 0,06 | 0,18 |
|  | **Ja** | 0,10 | 0,03 | 0,04 | 0,16 |
| Regionaal aanbod deeltijdse arbeid |  |  |  |  |  |
| Min | **15,47%** | 0,11 | 0,03 | 0,06 | 0,17 |
| Q1 | **18,18%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Med | **19,32%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Q3 | **20,40%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Max | **25,5%** | 0,14 | 0,03 | 0,07 | 0,20 |
| Proportie kleuters |  |  |  |  |  |
| Min | **7,13%** | 0,11 | 0,03 | 0,06 | 0,16 |
| Q1 | **10,12%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Med | **10,18%** | 0,12 | 0,03 | 0,06 | 0,18 |
| Q3 | **11,63%** | 0,13 | 0,03 | 0,06 | 0,19 |
| Max | **15,19%** | 0,14 | 0,04 | 0,07 | 0,21 |
| Tijd sinds eerste geboorte | **2** | 0,15 | 0,04 | 0,07 | 0,22 |
|  | **3** | 0,15 | 0,04 | 0,08 | 0,22 |
|  | **4** | 0,11 | 0,03 | 0,06 | 0,16 |
|  | **5** | 0,08 | 0,02 | 0,04 | 0,11 |
|  | **6** | 0,06 | 0,01 | 0,03 | 0,08 |

Alle resultaten zijn significant met p < 0,005; behalve voor leeftijd bij eerste kind = 46 jaar (vet, cursief en onderlijnd), dit resultaat is niet significant en leeftijd bij eerste kind = 16 (aangeduid met \*), waarvoor geldt dat p<0,05

## 7. Discussie

Naar aanleiding van o.a. de theoretische voorspellingen van Easterlin (1975) werd gecontroleerd voor mogelijk endogeniteit van kinderopvang, waarbij niet-geobserveerde kenmerken een invloed hebben op zowel de keuze voor kinderopvang als de keuze voor een tweede kind. Eerder werd al aangehaald dat gezien de *Wald test of exogeneity* blijkt dat het voor elk van de drie modelspecificaties onwaarschijnlijk is dat het gebruik van kinderopvang exogeen is.

Uit de vergelijking van de resultaten uit de normale probitregressies en de regressies aan de hand van instrumentele variabelen blijkt bovendien dat de geschatte effecten uit deze laatste merkbaar robuuster zijn voor verschillende modelspecificaties. Het effect in de normale regressies is significant verschillend tussen het eerste model en de twee anderen: 0,16 in het eerste model, met een 95%-betrouwbaarheidsinterval gaande van 0,14 tot 0,18 en 0,063 in de twee andere modellen, met als grenzen van het 95%-betrouwbaarheidsinterval 0,039 en 0,087. Hoewel het geschatte effect van formele kinderopvang daalt van 0,65 in het eerste model tot 0,61 in het derde, ligt deze laatste waarde niet buiten het 95%-betrouwbaarheidsinterval van de eerste schatting (0,45 – 0,85). De schatter is daarom robuust voor toevoeging van individuele en regionale kenmerken, wat een indicatie is voor de kwaliteit van de gebruikte instrumenten.

De resultaten van het eerste stadium van de schattingen aan de hand van instrumentele variabelen (zie *tabel 7 in de appendix (C))* tonen dat voor twee van de drie instrumenten, de theoretisch voorspelde relatie bevestigd wordt. Een groter aantal volwassenen is een aanwijzing dat het gezin toegang heeft tot substituten voor formele kinderopvang, vandaar de negatieve relatie. Een negatieve perceptie van de lokale opvangmogelijkheden hangt ook negatief samen met het gebruik ervan.

De proportie gesubsidieerde kinderopvang in een gemeente is verrassend negatief gecorreleerd met het gebruik van formele kinderopvang voor het eerste kind. Er werd een positieve relatie voorspeld omdat een grotere proportie gesubsidieerde opvang zou betekenen dat de gemiddelde prijs van kinderopvang in die gemeente lager is. Een mogelijke verklaring voor deze negatieve correlatie is dat, in gemeenten waar de vraag groter is dan het aanbod, de niet-gesubsidieerde opvang op een efficiëntere manier kan inspelen op die vraag. Dit zou echter betekenen dat deze variabele zelf endogeen is aan vruchtbaarheidsbeslissingen, wat twijfels doet rijzen over het gebruik ervan als instrument.

Voor ontevredenheid met de plaatselijke opvangvoorzieningen kan dan weer geargumenteerd worden dat een negatieve perceptie van het lokale aanbod aan kinderopvang een rechtstreeks effect heeft op de keuze voor een tweede kind. Gezinnen met een bescheiden inkomen, maar met twee werkende partners die aangewezen zijn op formele kinderopvang kunnen bijvoorbeeld beslissen om geen tweede kind te nemen wanneer opvang zeer duur is. Dit kan leiden tot een negatief beeld over de plaatselijke opvang. Daarnaast wil meer dan twee volwassenen in een huishouden niet automatisch betekenen dat die extra aanwezige volwassene(n) een substituut zijn voor formele kinderopvang.

Om de validiteit en relevantie van de gebruikte instrumenten uitgebreider te testen werd een alternatief model geschat dat deze testen toelaat. Aangezien ze slechts indicatief zijn voor de kwaliteit van de instrumenten in de gebruikte probit-schattingen, werden de resultaten opgenomen in de appendix (B), waar ze ook uitgebreid besproken worden. Voor zover de resultaten een goede indicatie vormen tonen ze aan dat er geen problemen zijn met de gebruikte instrumenten.

Twee conclusies kunnen uit de resultaten getrokken worden over de aard van de enodgeniteit. Ten eerste blijkt het effect van kinderopvang onderschat te worden wanneer hiervoor niet gecontroleerd wordt. Deze neerwaartse bias is ofwel te wijten aan niet-geobserveerde variabelen die positief gecorreleerd zijn met het gebruik van kinderopvang, maar negatief met de kans op een tweede geboorte. Ofwel gaat het om niet-geobserveerde variabelen die negatief correleren met het gebruik van kinderopvang en positief met de kans op een tweede geboorte. Een andere mogelijkheid is een combinatie van de twee.

Een tweede, maar gezien het beperkt aantal opgenomen regionale variabelen voorzichtige conclusie over de aard van de endogeniteit is dat het in dit onderzoek vooral gaat om individuele niet-geobserveerde heterogeniteit. Toevoeging van individuele kenmerken zorgt er namelijk voor dat de parameter zoals die geschat is in de normale probitmodellen meer dan halveert. Toevoeging van de regionale variabelen verandert deze schatter echter niet.

De resultaten uit de drie modellen met instrumentele variabelen tonen een positief verband tussen het gebruik van kinderopvang voor het eerste kind op de kans op een tweede geboorte. Dit ligt in lijn met de theoretische verwachting dat het gebruik van kinderopvang de opportuniteitskosten van kinderen doet dalen (Becker, 1981) en de bevindingen van eerder empirisch onderzoek (Baizàn, 2009; Rindfuss et al. 2010). De eerste hypothese wordt hiermee bevestigd. Aangezien instrumentele variabelen toelaten om causale inferenties te doen, kan op basis van het meest uitgebreide model gesteld worden dat het gebruik van kinderopvang voor het eerste kind de kans op een tweede kind jaarlijks met ongeveer 12% verhoogt.

Geen van de toegevoegde individuele kenmerken geeft onverwachte resultaten, met uitzondering van het niet-significante effect van de herkomstvariabele. Een waarschijnlijke verklaring hiervoor is dat het begrip *autochtoon* te breed is om überhaupt uitspraken te doen over herkomst en mogelijke culturele verschillen.

Toevoeging van de regionale variabele die de proportie kleuters per gemeente voorstelt in 2001, leert dat ook socialiserende factoren een rol spelen in de keuze voor kinderen (Lesthaeghe & Surkyn, 1988; Kohler, 2001 volgens Bongaarts 2002). Wanneer de proportie kleuters groter is in een bepaalde gemeente is dat namelijk een indicatie dat er 1) meer geboorten zijn[[8]](#footnote-8) en 2) het gaat om een meer “kindvriendelijke” gemeente.

Beide interpretaties zijn echter problematisch. Het is niet ondenkbaar dat de relatie tussen vruchtbaarheid en regionale karakteristieken in grote mate endogeen is. Zowel de eerste interpretatie duidt op een mogelijke oorzaak van endogeniteit (niet-geobserveerde heterogeniteit), alsook de tweede, met name zelfselectie, wanneer meer ouders met een voorkeur voor kinderen ervoor kiezen om in die gemeenten te gaan wonen en simultaneïteit wanneer een hogere vruchtbaarheid plaatselijke besturen ertoe aanzet om meer kindvriendelijke maatregelen te nemen. De robuustheid van het effect van formele kinderopvang na toevoeging van de regionale kenmerken betekent wel de bevestiging van de vierde hypothese.

De gerapporteerde marginale effecten uit *tabel 4,* die berekend zijn op basis van de derde modelspecificatie, laten toe een meer gedetailleerde beschrijving te geven van het effect van formele kinderopvang op de kans op een tweede geboorte. Deze effecten zijn berekend bij niet-gebruik van formele kinderopvang, waardoor de interpretatie van een eenheidsverandering overeenkomt met het verschil tussen niet en wel gebruik maken van formele kinderopvang.

Een eerste inspectie van de resultaten leert dat, in lijn met de theoretische verwachtingen, het gebruik van formele kinderopvang overal een positief effect heeft, zowel wat de geschatte marginale effecten betreft als de 95%-betrouwbaarheidsintervallen. De enige uitzondering hierop is de ondergrens voor het betrouwbaarheidsinterval bij het geschatte marginale effect van kinderopvang voor een vrouw die 46 jaar was bij de geboorte van het eerste kind (-0,02), maar dit effect is niet significant. De voor de hand liggende verklaring is hier dat de kans op een geboorte sowieso al klein is op die leeftijd, zowel om biologische redenen (de meeste vrouwen ervaren de menopauze tussen de 45 en de 55 jaar), maar ook omdat het risico op afwijkingen van het kind groter wordt naarmate de leeftijd van de moeder, wat een negatief effect heeft op de vruchtbaarheidsbeslissingen van oudere vrouwen.

Het marginale effect is groter voor vrouwen die hoger opgeleid zijn (15%) dan voor lager opgeleide vrouwen (9%) en vrouwen met een middelbare opleiding (10%). Aangezien de opportuniteitskosten van hoger opgeleide vrouwen waarschijnlijk hoger, is dit in lijn met de theoretische voorspellingen. Hoger opgeleide vrouwen hebben er dan namelijk meer baat bij om te participeren op de arbeidsmarkt, wat mogelijk gemaakt wordt door gebruik te maken van kinderopvang. Het faciliteren van de combinatie werk-gezin zorgt er dan voor dat de kans op een tweede geboorte stijgt. De tweede hypothese (b) wordt hiermee bevestigd.

Hetzelfde kan niet gezegd worden over de tewerkstellingskansen. Het marginale effect van het gebruik van formele kinderopvang is hier niet significant verschillend tussen de vrouwen met hogere en lagere kansen op de arbeidsmarkt en lijkt bovendien op het eerste gezicht dalend te zijn (van 14% voor vrouwen met minimale kans op tewerkstelling tot 12% voor vrouwen met een maximale kans). Hypothese 2(a) wordt hierdoor niet zonder meer bevestigd.

De verklaring hiervoor ligt waarschijnlijk bij de conceptie van deze indicator. Tewerkstellingskansen werden berekend omdat het onzeker is dat de arbeidsmarktstatus op het moment van de census (2001) dezelfde is als in de daaropvolgende jaren (2002-2005). Om de reële arbeidsmarktstatus van de vrouwen te benaderen werd daarom op basis van de censusgegevens een longitudinale variant opgesteld die deze kansen meet op basis van leeftijd, opleidingsniveau, herkomst en de gemeentelijke proportie deeltijds werk. Deze indicator zegt daarom weinig over de arbeidsmarktkansen van vrouwen en moet eerder bekeken worden als “de kans dat een vrouw met één kind aan het werk is”. Aangezien de variabelen die gebruikt werden om deze kansen te berekenen (met uitzondering van leeftijd) ook meegenomen werden in de uiteindelijke regressies wordt het geschatte effect onbetrouwbaar, wat ook blijkt uit de geschatte coëfficiënten in model 2 en 3, waarbij deze na toevoeging van de regionale kenmerken meer dan halveert en bovendien niet langer significant is.

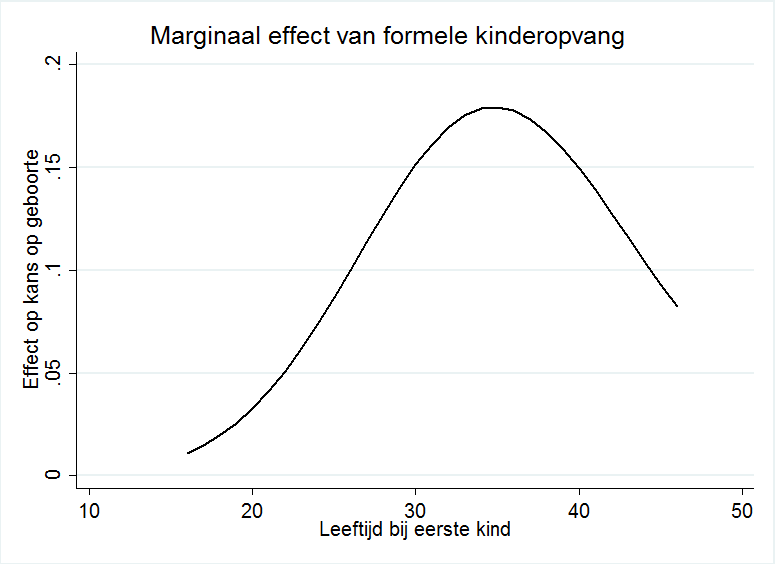
Wat wel in de richting van de theoretische voorspellingen over tewerkstellingskansen wijst is dat het marginale effect van kinderopvang groter wordt naarmate de regionale proportie deeltijds werk stijgt. Aangezien vrouwen vaker deeltijds werken is een groter aanbod aan deeltijds werk een indicatie van beter arbeidsmarktopportuniteiten voor vrouwen. Hoewel de marginale effecten van kinderopvang stijgen met de proportie deeltijds werk (van 11% naar 14%), zijn ze niet significant verschillend van elkaar.

Een tweede aanwijzing volgt uit de vergelijking van de kans dat een vrouw met één kind tewerkgesteld is, naar leeftijd met een alternatief model – resultaten niet opgenomen – waarbij leeftijd bij de geboorte van het eerste kind vervangen werd door leeftijd an sich. De berekende marginale effecten van formele kinderopvang tonen hier dat het effect het grootst is (12%, *p<0,005*) voor vrouwen van 33 jaar. De kans op tewerkstelling is het grootst voor vrouwen van 31 jaar.

Meer gedetailleerde inspectie van de marginale effecten zoals die uit *tabel 4* leert bovendien dat het effect van formele kinderopvang het grootst is voor vrouwen die hun eerste kind kregen op een leeftijd van 35 jaar (18%, *p < 0,01*). Dit wordt geïllustreerd in *figuur 7*, waarin het effect van kinderopvang op de kans op een tweede geboorte wordt voorgesteld in functie van de leeftijd bij het eerste kind.

Samen genomen zijn al deze resultaten een indicatie dat formele kinderopvang vooral een rol speelt in de vruchtbaarheidsbeslissingen van vrouwen die eerst hun arbeidsmarktpositie consolideren en daarom het eerste kind uitstellen. Gezien de problemen met de gebruikte indicator en het feit dat ook tewerkstelling waarschijnlijk endogeen is aan vruchtbaarheidsbeslissingen is verder onderzoek aangewezen.

Figuur : Marginaal effect van gebruik van formele kinderopvang naargelang leeftijd bij geboorte van het eerste kind



Bron: Census 2001, eigen bewerking

Wat verder opvalt en geïllustreerd wordt in *figuur 7* is dat het marginaal effect van het gebruik van formele kinderopvang afneemt in functie van de tijd sinds de geboorte van het eerste kind. Ook dit is geen verrassende vaststelling en ligt in lijn met de verwachtingen, aangezien de kans op een geboorte sowieso afneemt naarmate de tijd sinds het eerste kind langer wordt.

Figuur : Marginaal effect van formele kinderopvang naargelang tijd sinds geboorte eerste kind (in jaren)



Bron: Census 2001, eigen bewerking

Een laatste vaststelling is dat het marginale effect van formele kinderopvang niet significant verschilt tussen vrouwen die gecombineerd gebruik maken van zowel formele als informele opvang (13% tegenover 11%). Het effect lijkt bovendien stijgend. De derde hypothese, die stelde dat aangezien informele en formele kinderopvang substituten zijn, het effect van deze laatste zou dalen, wordt hiermee niet bevestigd.

Het effect van informele kinderopvang zelf is significant in model 2 en marginaal significant in model 3. Mogelijk tonen de resultaten dat er een intensiteitseffect is voor het gebruik van opvang. De data liet echter niet toe om de precieze intensiteit van zowel formele als informele kinderopvang te meten. De schijnbare stijging van het marginale effect bij gebruik van informele kinderopvang is daarom aannemelijk voor zover het gezamenlijke gebruik van formele en informele opvang effectief betekent dat vaker gebruik gemaakt wordt van kinderopvang.

Tot slot van deze bespreking is er een belangrijke kanttekening te plaatsen bij de efficiëntie van de resultaten. Hoewel de meeste gerapporteerde marginale effecten in lijn liggen met theoretisch voorspeld werd, zijn de betrouwbaarheidsintervallen van dien aard dat het statistisch mogelijk is dat geen enkele ervan afwijkt van het gemiddeld marginaal effect van het gebruik van formele kinderopvang van 12%. Dit is een rechtstreeks gevolg van het gebruik van instrumentele variabelen, die weliswaar consistent, maar minder efficiënt zijn.

## 8. Beperkingen van het onderzoek

Bij al deze resultaten zijn enkele belangrijke kritische bedenkingen te maken. Hierbij wordt het onderscheid gemaakt tussen het onderzoeksopzet en de gebruikte data.

### 8.1 Opzet

Het opzet van het onderzoek laat ten eerste niet toe dat vrouwen gevolgd worden overheen al hun vruchtbare jaren. Vrouwen met een eerste kind in 1999 worden bijvoorbeeld maximaal zes jaar opgevolgd. Hoewel de kans op een tweede kind verkleint in functie van de tijd sinds een eerste geboorte, is het evident dat ook na zes jaar nog een tweede kind geboren kan worden.

Een tweede opmerking over het opzet is dat geen rekening gehouden wordt met ongeplande zwangerschappen. Buysse et al. (2013) kwamen tot de conclusie dat ongeveer 25% van de zwangerschappen tussen 2000 en 2011 ongepland was, terwijl slechts 5% van de zwangerschappen eindigde met een abortus. Dit plaatst vraagtekens bij de theorie van perfecte planning (Bumpass & Westoff, 1970 volgens Bongaarts & Westoff, 2000). Er zou echter geargumenteerd kunnen worden dat, wanneer de optie om te kiezen voor een abortus aanwezig is, een ongeplande zwangerschap uit te dragen ook een keuze is, en daarom past bij perfecte planning. Dit is echter voer voor een discussie waar minstens een extra thesis mee gevuld kan worden.

Daarnaast zorgt het probleem van selectiviteit van de steekproef ervoor dat uiterst omzichtig omgesprongen moet worden met de externe validiteit van de resultaten. Aangezien enkel vrouwen opgenomen werden die reeds een eerste kind hadden, zijn deze resultaten niet zonder meen te veralgemenen naar de hele populatie. Bovendien spelen er andere factoren bij verschillende pariteitsprogressies.

### 8.2 Data

Naast de problemen met het opzet is het enigszins problematisch dat de beschikbare data grotendeels cross-sectioneel van aard is. Uitgezonderd de geboorte van een tweede kind, leeftijd van de moeder en de geschatte tewerkstellingskansen (die een functie zijn van onder andere de leeftijd), zijn er geen variabelen met consecutieve waarnemingen. In het geval van tewerkstelling zijn de geschatte kansen bovendien niet te interpreteren als arbeidsmarktopportuniteiten maar simpelweg als “de kans dat een vrouw met één kind aan het werk is, gegeven bepaalde kenmerken”. Desondanks kan aangenomen worden dat ze de reële arbeidsmarktpositie in de jaren volgend op de census beter benaderen dan een dummyvariabele die enkel de status op het moment van de census meet.

De cross-sectionele natuur van de data zorgt er wel voor dat potentieel belangrijke variabelen (burgerlijke staat, tewerkstelling van de partner, voltijds/deeltijds werk) niet zijn opgenomen in het model omdat de status van deze variabelen in de periode rond een tweede geboorte onzeker is. Het is bovendien niet ondenkbaar dat het feit dat een deel van de vrouwen die in de geobserveerde periode geen tweede kind kreeg te wijten kan zijn aan veranderingen in de relatiestatus (echtscheiding, einde cohabitatie,…).

Een tweede bedenking is de ruwe selectie die gemaakt is binnen de data. Het beperkte opzet van dit onderzoek liet niet toe om potentieel problematische data gedetailleerd te onderzoeken, waardoor de keuze gemaakt is om deze observaties te verwijderen. Een voorbeeld is het eerder aangehaalde probleem met niet-biologische kinderen. Hoewel de selectie van de data grotendeels uitsluit dat er meer dan één kind aanwezig is in het gezin, is dit geen absolute garantie. Het is mogelijk de data te controleren op verwantschappen, maar dit was niet realistisch binnen het tijdsbestek van dit onderzoek.

Daarnaast zijn er problemen met de verdeling van enkele van de variabelen. Tewerkstellingskans is bijzonder linksscheef verdeeld (*figuur 8*) net als de gemeentelijke proportie gesubsidieerde kinderopvang (*figuur 9*). De concentratie van waarnemingen bij hogere waarden van deze twee variabelen vertekent mogelijk de geschatte effecten. Voorts geeft slechts 0.72% van de respondenten aan ernstige gezondheidsproblemen te hebben, al gaat het nog steeds om 234 respondenten. De problemen met het concept “autochtoon” werden eerder ook besproken.

|  |  |
| --- | --- |
| Figuur : verdeling tewerkstellingskansen    Bron: Census 2001, eigen bewerking | Figuur : Verdeling proportie gesubsidieerde opvang    Bron: Census 2001, eigen bewerking |

## 9. Conclusies

Op basis van data uit de Belgische census van 2001, gecombineerd met registerdata en data van Kind en Gezin, werd in dit onderzoek het effect van het gebruik van formele kinderopvang voor een eerste kind op de kans op het krijgen van een tweede kind voor vrouwen in Vlaanderen onderzocht. De resultaten van probitregressies met instrumentele variabelen maken het mogelijk te controleren voor endogeniteit. Niet controleren voor endogeniteit blijkt te zorgen voor onderschatting van het effect, wat met het oog op gezinsbeleid tot verkeerde evaluaties en beslissingen zou kunnen leiden. Deze thesis is de eerste keer dat de techniek van instrumentele variabelen in dit soort setting gebruikt wordt en ondanks enkele kleine bemerkingen presteren ze voortreffelijk.

Uit de analyses blijkt dat vrouwen die gebruik maken van formele kinderopvang voor hun eerste kind jaarlijks 12% meer kans hebben op een tweede geboorte. Daarnaast werd bevestigd dat het effect groter is voor hoger opgeleide vrouwen. Voor vrouwen met een hogere kans op tewerkstelling werden slechts indicaties gevonden van een groter effect van het gebruik van formele kinderopvang op de kans op een tweede geboorte. Gelijktijdig gebruik van informele kinderopvang heeft dan weer geen invloed op het effect van formele kinderopvang. De resultaten zijn robuust voor modelspecificaties waarin rekening gehouden werd met regionale kenmerken van de arbeidsmarkt en de mate waarin een gemeente “kindvriendelijker” is.

De resultaten stemmen overeen met theoretische literatuur over opportuniteitskosten (Becker, 1981), aangezien vrouwen met hogere tewerkstellingskansen minder kans hebben op een tweede geboorte. Voor hen is de opportuniteitskost van het krijgen van kinderen namelijk groter. Anderzijds is opleidingsniveau positief gecorreleerd aan de kans op een tweede geboorte, wat wijst op een inkomenseffect. Controle aan de hand van instrumentele variabelen wijst op het bestaan van endogene voorkeuren (Easterlin, 1975). De positieve relatie tussen de proportie kleuters in een gemeente toont dan weer dat sociale context (Lesthaeghe & Surkyn, 1988) in verband staat met individuele vruchtbaarheidsbeslissingen. Ze liggen hiermee in lijn met eerder onderzoek dat controleerde naar de endogeniteit van kinderopvang (Baizàn, 2009; Rindfuss, 2010).

Analyses aan de hand van probitregressies met instrumentele variabelen blijken een valabel alternatief voor de gebruikelijke hazard-modellen op basis van logistische regressie met *fixed effects*. Op deze manier is de gebruikte techniek een oplossing voor onderzoekers die voornamelijk beschikken over cross-sectionele in plaats van longitudinale data, maar biedt ze evenzeer een alternatief wanneer de assumptie van in de tijd onveranderlijke niet-geobserveerde kenmerken dubieus is.

Alles tezamen zijn de resultaten uit deze thesis een belangrijke mogelijke nieuwe richting voor het onderzoek naar de effecten van overheidsbeleid op vruchtbaarheid. Verder onderzoek is echter nog nodig om het precieze effect voor alle geboortepariteiten te onderzoeken om zo uitspraken te kunnen doen over het effect van kinderopvang op het totale geboortecijfer.

## 10. Toekomstig onderzoek

Hoewel de resultaten geen uitsluitsel geven over de algemene invloed van het gebruik van formele kinderopvang op het geboortecijfer, zijn ze een belangrijke indicatie van een positief effect. Kinderopvang speelt een belangrijke rol in het vergemakkelijken van de combinatie werk-gezin. Om het totale effect op het geboortecijfer te berekenen is het voor toekomstig onderzoek interessant om (in navolging van Rindfuss et al, 2010), de analyse te herhalen voor gezinsvorming (eerste geboorte) enerzijds en hogere geboortepariteiten anderzijds. Op die manier kunnen aan de hand van microsimulaties uitspraken gedaan worden over het effect van formele kinderopvang op het totaal aantal geboorten.

Daarnaast is bijkomend onderzoek nodig naar het verband tussen tewerkstelling en uptake van kinderopvang. In dit onderzoek werd aangenomen dat de beslissing om te werken voorafgaat aan de beslissing om gebruik te maken van formele kinderopvang en de keuze voor een tweede kind. Het is echter mogelijk dat die beslissing simultaan genomen wordt.

Hoewel het gebruik van instrumentele variabelen een vaak gebruikte techniek is om te controleren voor endogeniteitsproblemen op te lossen, is het lang niet de enige methode. Eerder onderzoek door Baizàn (2009) en Rindfuss et al. (2010) gebruikten bijvoorbeeld *fixed effects*-modellen. De nadelen van beiden zijn reeds besproken. Voor instrumentele variabelen is het belangrijkste nadeel het vinden van goede instrumenten, voor *fixed effects*-modellen is de assumptie dat de niet geobserveerde variabelen onveranderlijk zijn doorheen de tijd.

Recent werd door Bartus (2012) voorgesteld om in dit soort onderzoek *multilevel multiproces*-modellen te gebruiken. Het belangrijkste voordeel van deze techniek lijkt dat ze de onderzoeker in staat stelt om bij de modelspecificatie een onderscheid te maken tussen drie verschijningsvormen van het endogeniteitsprbleem: 1) omgekeerde causaliteit en simultane keuzes, 2) niet-geobserveerde heterogeniteit en 3) problemen van zelfselectie.

Een volledig begrip van de werking van dit soort modellen en een diepgaande analyse van de geschatte effecten liggen buiten de scope van deze paper, maar het lijkt desalniettemin een interessante piste voor meer ervaren onderzoekers. Als voorzichtige aanzet werd een model geschat dat vergelijkbaar is met Model 1, maar waarbij het endogeniteitsprobleem werd opgevat als een probleem van zelfselectie. In zijn simpelste vorm komt de methode neer op een simultane schatting van twee probit-vergelijkingen, een eerste voor de kans op het gebruik van formele kinderopvang (op basis van de drie variabelen die voorheen als instrumenten gebruikt werden) en tweede vergelijking voor de kans op een tweede geboorte.

Het resultaat is een schatting van het gemiddelde marginale effect van formele kinderopvang van 6,5% (*p<0,01)*. Hoewel deze schatting lager is dan deze van de gebruikte modellen in deze onderzoekspaper, is ze significant hoger dan de probitmodellen waarbij geen rekening gehouden werd met mogelijke endogeniteit van formele kinderopvang. Dit resultaat geeft daarom minstens een extra indicatie dat het effect van formele kinderopvang onderschat wordt wanneer niet gecontroleerd wordt voor endogeniteit. De grootte van deze onderschatting en concreter onderzoek van het mogelijke nut van dit soort modellen levert interessante stof voor verder onderzoek.

## Appendix

### A. Probit vs Logit vs LPM

Om een antwoord te bieden op mogelijke bedenkingen over de geschiktheid van het probitmodel voor dit type analyse wordt in dit onderdeel een pragmatisch overzicht gegeven van enkele argumenten. Hierbij wordt vertrokken van het vaak gebruikte *discrete-time* hazardmodel dat traditioneel van toepassing zou zijn op de gebruikte data in dit onderzoek, gezien het beperkt aantal punten waarop respondenten geobserveerd worden (jaarlijks, maximaal 6 jaar). Formeel ziet dit er als volgt uit:

 (9)

Waarbij het linkerlid het natuurlijke logaritme is van de odds dat er een geboorte plaats vindt bij persoon *i* in periode *t*.  staat voor een set constanten voor elk tijdsinterval (t=1,2,…), een vector met geschatte parameters en stelt onafhankelijke variabelen voor die verschillende waarden kunnen aannemen op verschillende tijdstippen.

Daarnaast werd een eenvoudig *linear probability model* (LPM) geschat om bijkomende informatie te krijgen:



In principe is dit type modellen weinig geschikt om binaire keuzes te analyseren aangezien de eruit

voorspelde kansen buiten het interpreteerbare [0,1]-interval kunnen vallen. Voor de analyse van gemiddelde marginale effecten is dit echter van weinig belang en geven ze een goede benadering van niet-lineaire modellen als probit en logit (Angrist & Pischke, 2009).

De marginale effecten uit elk van de modellen en voor alle drie de modelspecificaties worden weergegeven in *tabel 5*. Uit deze tabel blijkt dat de schattingen in zeer grote mate consequent zijn overheen de verschillende modelspecificaties. Dit is een indicatie dat er weinig verschil is tussen de drie schattingsmethoden, zo lang het gaat over de gemiddelde marginale effecten. Dit is een belangrijke verantwoording voor het gebruik van probit-modellen, die dan weer als voordeel hebben dat er met instrumentele variabelen gewerkt kan worden.

Tabel : Resultaten voor schattingen a.d.h.v. verschillende technieken en modelspecificaties

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | | Model 3 | | |
|  | **LPM** | **Logit** | **Probit** | **LPM** | **Logit** | **Probit** | **LPM** | **Logit** | **Probit** |
| dy/dx | 0,046  (0,003) | 0,041  (0,002) | 0,041  (0,002) | 0,017  (0,003) | 0,016  (0,003) | 0,017  (0,003) | 0,016  (0,003) | 0,016  (0,003) | 0,017  (0,003) |

dy/dx: marginale verandering in de kans op een tweede geboorte bij een verandering van formele kinderopvang van 0 naar 1.   
Model 1: enkel tijdsindicatoren en formele opvang; Model 2: uitbreiding met: informele opvang, tewerkstellingskans, opleidingsniveau, leeftijd bij eerste geboorte (lineair en kwadratisch), gezondheidsstatus herkomst; Model 3: uitbreiding metproportie geboorten, proportie deeltijds werk, proportie kleuters; standaardfouten tussen haakjes. Alle effecten zijn significant met p<0,005.

### B. Validiteit en relevantie van de gebruikte instrumenten

In de vorige paragraaf van deze appendix werd reeds aangetoond dat op het vlak van marginale effecten er weinig verschil is tussen een LPM en een probit-model. Een nadeel van deze laatste is dat er geen eenvoudige technieken beschikbaar zijn om de validiteit en de relevantie van de gebruikte instrumenten formeel te testen. Deze bestaan wel voor lineaire modellen.

Om extra indicaties te bekomen van de kwaliteit van de gebruikte instrumenten werden de regressies met instrumentele variabelen daarom nogmaals uitgevoerd aan de hand van het lineaire model. De resultaten van de parameters zijn hier niet belangrijk (hoewel ze voor het gemiddelde marginale effect van formele kinderopvang niet significant verschillen van die uit de probitmodellen), wat wel belangrijk is zijn de teststatistieken voor de gebruikte variabelen. Deze worden weergegeven in *tabel 6.*

Tabel : Robuustheidschecks voor gebruikte instrumenten a.d.h.v. lineaire regressie

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 1 | Model 2 | Model 3 |
| Kleibergen-Paap rk LM statistic Chi² (df) | 641,7\*\*\*  (3) | 249,51\*\*\*  (3) | 248,17\*\*\*  (3) |
| Kleibergen-Paap rk Wald F statistic | 215,44\* | 83,390 | 82,938 |
|  |  |  |  |
| Hansen J Chi² (df)  *p-waarde* | 5,158 (2)  *0,0759* | 0,448  *0,7993* | 0,355  *0,8374* |

Model 1: enkel tijdsindicatoren en formele opvang; Model 2: uitbreiding met: informele opvang, tewerkstellingskans, opleidingsniveau, leeftijd bij eerste geboorte (lineair en kwadratisch), gezondheidsstatus herkomst; Model 3: uitbreiding met proportie geboorten, proportie deeltijds werk, proportie kleuters. Stock-Yogo weak ID-test 5% kritische waarde: 13,91  
\*p<0,05 \*\*\*p<0,005)

De Kleibergen-Paap LM-Chi²-waarde is een testwaarde voor de identificatie van het model. De nulhypothese is dat het model ondergeïdentificeerd is. Wanneer deze hypothese verworpen wordt is dat een indicatie dat de gebruikte instrumenten relevant zijn, waarmee bedoeld wordt dat ze gecorreleerd zijn met de endogene variabele. De nulhypothese wordt in elk van de drie modellen verworpen.

De Kleibergen-Paap Wald F-statistiek is een F-test om te kijken of de gebruikte instrumenten zwak zijn. Onder de nulhypothese correleren de gebruikte instrumenten maar weinig met de endogene regressor. De Stock-Yogo-kritische waarde voor een significantieniveau van 5% is 13,91. In elk van de drie modellen zit de F-statistiek ruim boven die waarde, waardoor de nulhypothese (dat de instrumenten zwak zijn) verworpen kan worden.

Tot slot toetst de Hansen-J-test de gezamenlijke nulhypothese dat de gebruikte instrumenten niet gecorreleerd zijn met de foutenterm en dat ze terecht niet zijn opgenomen als aparte variabelen (en enkel als instrument dienen). Anders dan bij de twee vorige hypothesetoetsen is het hier dus de bedoeling dat de nulhypothese niet verworpen kan worden. Voor de eerste modelspecificatie zien we dat de nulhypothese enkel verworpen kan worden met marginale significantie van *p<0,1*. Voor de twee andere modelspecificaties kan de nulhypothese niet verworpen worden.

Tot slot is het vermeldenswaardig dat aparte schattingen van de *redundancy* van de verschillende instrumenten –resultaten niet opgenomen– leren dat geen enkele van de gebruikte instrumenten overbodig is.

### C. Eerste stadium IV-regressies (probit)

Tabel : Resultaten uit eerste stadium van IV-regressies, endogene variabele = gebruik van formele kinderopvang voor het eerste kind

|  |  |  |  |
| --- | --- | --- | --- |
| Covariaten | Model 1  (IVprobit) | Model 2 (IVprobit) | Model 3  (IVprobit) |
| Tijdsindicatoren late entry (ref = 2) |  |  |  |
| 3 | -0,025\*\*\*  (0,005) | -0,002  (0,004) | -0,002  (0,004) |
| 4 | -0,047\*\*\*  (0,005) | -0,008+  (0,004) | -0,008+  (0,004) |
| 5 | -0,055\*\*\*  (0,005) | -0,011\*  (0,004) | -0,011\*  (0,004) |
| 6 | -0,064\*\*\*  (0,007) | -0,013\*  (0,006) | -0,012\*  (0,006) |
| constante | 0,85\*\*\*  (0,011) | 0,092+  (0,052) | 0,050  (0,054) |
| Instrumenten |  |  |  |
| Aantal volwassenen in het gezin | -0,10\*\*\*  (0,013) | -0,066\*\*\*  (0,011) | -0,066\*\*\*  (0,011) |
| Negatieve perceptie van lokale kinderopvang (ref = nee) | -0,01\*\*\*  (0,004) | -0,015\*\*\*  (0,003) | -0,015\*\*\*  (0,003) |
| Gemeentelijke proportie gesubsidieerde kinderopvang | -0,29\*\*\*  (0,012) | -0,139\*\*\*  (0,010) | -0,148\*\*\*  (0,011) |
| Individuele kenmerken |  |  |  |
| Informele kinderopvang (ref=nee) |  | -0,519\*\*\*  (0,003) | -0,518\*\*\*  (0,003) |
| Leeftijd bij eerste kind |  | 0,035\*\*\*  (0,005) | 0,032\*\*\*  (0,005) |
| Leeftijd bij eerste kind (kwadraat) |  | -0,001\*\*\*  (0,0001) | -0,0005\*\*\*  (0,0001) |
| Opleidingsniveau (ref=lager) |  |  |  |
| Hoger |  | 0,178\*\*\*  (0,008) | 0,172\*\*\*  (0,008) |
| Midden |  | 0,053\*\*\*  (0,006) | 0,050\*\*\*  (0,006) |
| Tewerkstellingskans |  | 0,155\*\*\*  (0,039) | 0,187\*\*\*  (0,041) |
| Slechte gezondheid (ref=nee) |  | -0,217\*\*\*  (0,015) | -0,217 |
| Autochtoon (ref=nee) |  | 0,066\*\*\*  (0,009) | 0,062\*\*\*  (0,010) |
|  |  |  |  |
| Regionale kenmerken |  |  |  |
| Proportie geboorten |  |  | 8,126\*\*\*  (1,858) |
| Proportie kleuters |  |  | -0,298\*\*  (0,135) |
| Proportie deeltijds werk |  |  | 0,262\*\*\*  (0,262) |
|  |  |  |  |

Standaardfouten tussen haakjes; + p < 0,1 ; \* p<0,05 ; \*\* p<0,01 ; \*\*\*p<0,005

1. Een belangrijke noot hierbij is dat perfecte planning niet lijkt overeen te komen met de realiteit. Buysse et al. (2013) schatten op basis van een steekproef van 637 respondenten (mannen en vrouwen) dat ongeveer 25% van de zwangerschappen in Vlaanderen tussen 2000 en 2011 *ongepland* was, maar dat slechts 20% van deze zwangerschappen resulteerde in een abortus. Dit zou betekenen dat mogelijks 20% van de geboorten niet het gevolg zijn van een bewuste keuzestrategie. Een tegenargument is dat kiezen voor al dan niet een abortus ook een vorm van planning is. [↑](#footnote-ref-1)
2. Letterlijk: “Have childcare problems ever led you to decide to have fewer children than you had originally planned?” [↑](#footnote-ref-2)
3. De leerplicht in Zweden start officieel bij 7 jaar [↑](#footnote-ref-3)
4. De precieze leeftijd wordt niet gegeven in de paper, maar in Noorwegen is de schoolplichtige leeftijd 6 jaar. Vermoedelijk is de variabele dus gemeten aan de hand van kinderen van 0-5 jaar. [↑](#footnote-ref-4)
5. Niet-erkende kinderdagverblijven kunnen via een attest van Kind en Gezin bewijzen dat ze aan de wettelijke voorwaarde voldoen. [↑](#footnote-ref-5)
6. Regelmatig: minstens één onderbroken periode van minstens 5 uur per week voor niet-schoolgaande kinderen en minstens één maal per week voor kinderen tussen 2,5 en 3 jaar die volledig naar de kleuterschool gaan (Hedebouw, 2009; p.36) [↑](#footnote-ref-6)
7. Analyses gebeurden aan de hand van het *Stata-*softwarepakket:StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP. [↑](#footnote-ref-7)
8. Deze stelling wordt mee gestaafd door het feit dat in een alternatieve modelspecificatie – resultaten niet opgenomen – waarbij de proportie kleuters weggelaten werd, de variabele die de proportie geboorten in een gemeente in 2001 voorstelt wél significant is met: *p < 0,001*), wat zou kunnen wijzen op regionale normverschillen qua gewenst kinderaantal. [↑](#footnote-ref-8)