

Appendix: technische toelichting

Modelspecificatie

Om te onderzoeken welke factoren invloed hebben op de kennisintensiteit van de sectorstructuur gebruiken we het volgende algemene model:

$$S_{i,t} = \alpha_i + \beta_1(S_{i,t-1}) + \beta(x_{i,t-j}) + \sum_i D_{i,t} + \varepsilon_{i,t} \quad (1)$$

Hierin is S het sectorstructureffect voor land i en jaar t , j is een vector die de vertraging van de variabelen aangeeft, α is een constante, x is een vector van onafhankelijke variabelen, β is een vector van coëfficiënten, D staat voor land- en tijdspecifieke effecten (*fixed effects*) en ε is een foutterm.

Er doen zich bij het schatten van dit eenvoudige model echter twee econometrische problemen voor. Ten eerste zijn onze variabelen in niveaus niet-stationair en is er een reëel risico op het schatten van zogenoemde nonsens-correlaties ([Granger en Newbold, 1974](#)). Een tweede econometrisch probleem is dat de coëfficiënten van een panelmodel met fixed effects en een vertraagde endogene variabele vertekend zijn (zie [Nickell, 1981](#)).

Dynamic OLS

Een manier om te corrigeren voor non-stationariteit van variabelen is om het model te schatten in eerste verschillen, maar dan verliezen we ook informatie over de langetermijnrelatie tussen de variabelen. Wanneer echter sprake is van zogenoemde *co-integratie* kunnen niet-stationaire variabelen wel in niveaus worden geschat. Co-integratie betekent dat de residuen van een combinatie van niet-stationaire variabelen wel stationair zijn over tijd. Om dit te testen maken we gebruik van de co-integratietest voor paneldata ontwikkeld door [Kao \(1999\)](#). Uit deze test blijkt dat sprake is van co-integratie van onze variabelen. [Kao en Chiang \(2001\)](#) laten zien dat de Dynamic Ordinary Least Squares (DOLS) schatter bij gecointegreerde panelmodellen een veel nauwkeurigere schatter is dan OLS of fully-modified OLS (FMOLS). Het DOLS model heeft de volgende specificatie:

$$S_{i,t} = \alpha_i + \beta(x_{i,t-j}) + \sum_{j=-q}^q \Delta x_{i,t+j} + \varepsilon_{i,t} \quad (2)$$

Het grote verschil van het DOLS-model ten opzichte van vergelijking (1) is het opnemen van vertragingen en versnellingen (lags en leads) van de onafhankelijke variabelen in eerste verschillen (Δ). Hiermee wordt gecorrigeerd voor autocorrelatie in de residuen ([Stock and Watson, 1993](#)). In dit geval hoeven we dus geen gebruik te maken van landen- en jaardummy's die normaal hiervoor corrigeren.

Robuustheidsanalyses: ECM en GMM

De DOLS-schatting heeft onze voorkeur, omdat we hiermee de langetermijnrelatie tussen het sectorstructureffect van landen en andere variabelen kunnen vaststellen. Echter, als robuustheidsanalyse voeren we ook schattingen uit met twee dynamische panelmodellen. In beide dynamische modellen maken we gebruik van *white errors* om te corrigeren voor heteroskedasticiteit in de residuen. Het eerste model is een foutcorrectiemodel (error-correction model: ECM), waarmee zowel het kortetermijn- als langetermijnverband integraal kan worden geschat. Er wordt ook een error-correctieparameter geschat (λ), die corrigeert voor kortetermijnenafwijkingen van variabelen van hun langetermijnevenwichtswaarde. Als er een positieve (negatieve) afwijking van

het langetermijnevenwicht is, wordt deze term dus negatief (positief), om zo de afwijking te verminderen. Het ECM-model wordt als volgt gespecificeerd:

$$\Delta S_{i,t} = \alpha_i + \beta_0 \Delta(S_{i,t-1}) + \beta \Delta(x_{i,t-j}) + \lambda[(S_{i,t-1}) - (\beta' x_{i,t-j}')] + \sum_i D_{i,t} + \varepsilon_{i,t} \quad (3)$$

In ons ECM-model nemen we ook een vertraagde endogene variabele mee om rekening te houden met autoregressieve processen. De sectorstructuur van een economie verandert immers weinig op korte termijn. Mogelijk is hierbij echter sprake van endogeniteit (Cameron en Trivedi, 2005), omdat de autoregressieve term gecorreleerd is met de foutterm. Als tweede robuustheidsanalyse schatten we daarom een eerste verschillen Generalised Method of Moments (GMM) model gebaseerd op [Arellano en Bond \(1991\)](#). Gebaseerd op dezelfde bron gebruiken we de vertraagde niveaus van onze afhankelijke en onafhankelijke variabelen als instrumenten voor de autoregressieve term $\Delta S_{i,t-1}$ en de onafhankelijke variabelen. Het GMM-model heeft de volgende vorm:

$$\Delta S_{i,t} = \beta_0 \Delta(S_{i,t-1}) + \beta \Delta(x_{i,t-j}^*) + \varepsilon_{i,t} \quad (4)$$

Waarbij x^* de vector van geïnstrumenteerde onafhankelijke variabelen is.

Data

We schatten bovenstaande drie modellen (vergelijking (2), (3) en (4)) voor een panel van 16 OESO-landen: België, Canada, Denemarken, Duitsland, Frankrijk, Griekenland, Italië, Japan, Noorwegen, Nederland, Oostenrijk, Portugal, Spanje, het Verenigd Koninkrijk, de Verenigde Staten en Zweden over een periode van drie decennia (1987-2017). Het sectorstructureffect en intrinsieke effect is berekend met data uit drie OESO-databases: de STAN-database, ANBERD-database en Main Science and Technology Indicators. Publieke R&D is ook berekend door gebruik te maken van data uit laatstgenoemde OESO-database. De arbeidskosten per eenheid product zijn berekend met data uit de OESO Economic Outlook, AMECO-database van de Europese Commissie, OESO Annual National Accounts, IMF World Outlook database en tot slot de Total Economy Database van de Conference Board. Voor de openheidsvariabele is data gebruikt uit de OESO Economic Outlook database.

Ierland en Finland

Landen waarvoor de data niet compleet zijn of veel sectoren ontbreken nemen we niet mee in onze analyse. Daarnaast hebben we er voor gekozen om Ierland en Finland niet mee te nemen in de panel-schattingen, vanwege aanzienlijke vertekeningen in de data.

In Ierland zijn door het aantrekkelijke belastingregime een flink aantal grote buitenlandse bedrijven gevestigd. Dit heeft een versturende werking op de data uit de nationale rekeningen die wij gebruiken. Zo dragen opbrengsten op basis van het gebruik van bedrijfsspecifiek intellectueel eigendom bij aan het Ierse bbp in plaats van dat van andere landen (zie [OECD, 2016](#)). Om een voorbeeld te geven: in Ierland is een aantal grote internationale farmaceutische bedrijven actief, zoals Pfizer, Johnson & Johnson en Roche, een sector waar traditioneel sprake is van een omvangrijke opbrengst uit licenties en patenten. Dit verklaart waarom de toegevoegde waarde van de Ierse farmaceutische industrie in sommige jaren met 30 procent is gegroeid. Doordat de farmaceutische sector ook nog erg R&D-intensief is, wordt de kennisintensiteit van de Ierse sectorstructuur in de nationale rekeningen flink overschat.

In Finland wordt de sectorstructuur historisch gedomineerd door de R&D-uitgaven van slechts één bedrijf: Nokia. Vroeger was Nokia een wereldwijde koploper op het gebied van mobiele technologie, maar met de opkomst van de smartphone in 2007 is de dominante positie van Nokia verdwenen. Tussen 2012 en 2014 daalden de R&D-uitgaven van Nokia van [vijf miljard euro naar twee miljard](#)

[euro](#). Ten opzichte van het totaal van R&D-uitgaven in Finland van ongeveer 7 miljard euro in 2012 is dit een grote daling en is er dus sprake van een trendbreuk.

Variabelen

Voordat we de schattingsresultaten van onze modellen bespreken, staan we stil bij de berekening van de variabelen in onze modellen. Het sectorstructureffect (S) wordt voor 16 landen als volgt berekend:

$$S_{i,t} = \sum_m \left(\frac{RD_{F,t,m}}{Y_{F,t,m}} \right) \times \left[\left(\frac{Y_{i,t,m}}{\sum_m Y_{i,t,m}} \right) - \left(\frac{Y_{F,t,m}}{\sum_m Y_{F,t,m}} \right) \right] \quad (5)$$

waarin RD staat voor de nominale R&D-uitgaven, Y voor de toegevoegde waarde en i voor het land en m voor de sector. De index F staat voor de buitenlandse benchmark. Het sectorstructureffect wordt dus berekend door de R&D-intensiteit (R&D-uitgaven als percentage van de toegevoegde waarde) van een bepaalde sector in de benchmark te vermenigvuldigen met het aandeel van die sector in de totale toegevoegde van het land zelf relatief ten opzichte van de benchmark.

Het intrinsieke effect (I) wordt berekend door:

$$I_{i,t} = \sum_i \left(\frac{Y_{i,t,m}}{\sum_m Y_{i,t,m}} \right) \times \left[\left(\frac{RD_{i,t,m}}{Y_{i,t,m}} \right) - \left(\frac{RD_{F,t,m}}{Y_{F,t,m}} \right) \right] \quad (6)$$

Bij het intrinsieke effect wordt dus louter gekeken naar de R&D-uitgaven van bedrijven in een sector ten opzichte van dezelfde sector in de benchmark. Zoals de term al zegt wordt er gekeken of bedrijven intrinsiek minder of meer uitgaven aan R&D vergeleken met hun buitenlandse concurrenten.

De publieke R&D-variabele wordt berekend door:

$$PUB_{i,t} = \left[\left(\frac{G_{i,t}}{Y_{i,t}} \right) - \left(\frac{G_{F,i}}{Y_F} \right) \right] + \left[\left(\frac{H_{i,t}}{Y_{i,t}} \right) - \left(\frac{H_{F,i}}{Y_F} \right) \right] \quad (7)$$

Waarbij G de R&D-investeringen zijn van publieke onderzoeksinstituten en H die van universiteiten. Deze variabele geeft dus weer in hoeverre een land meer investeert in publieke R&D dan de buitenlandse benchmark.

Voor het bepalen van de openheid van de economie gebruiken we de definitie van [Bassanini, Scarpetta en Hemmings \(2001\)](#):

$$O_{i,t} = \left(\frac{EX_{i,t}}{Y_{i,t}} \right) + \left[\left(1 - \frac{EX_{i,t}}{Y_{i,t}} \right) \times \left(\frac{IM_{i,t}}{Y_{i,t} + IM_{i,t} - EX_{i,t}} \right) \right] \quad (8)$$

Waarin EX staat voor de export van land i en IM voor de import van land i . De openheid van een economie wordt dus bepaald door de exportintensiteit van een economie (eerste term na gelijkteken) en de importpenetratie van de binnenlandse vraag (term tussen rechte haken). Omdat kleine landen per definitie meer blootstaan aan buitenlandse dan grote landen met een grote interne afzetmarkt, corrigeren we voor de omvang van de economie ([Donselaar, 2011](#)).

De arbeidskosten per eenheid product (ULC) worden bepaald door:

$$ULC_{i,t} = \left(\frac{\left(\frac{LC_{i,t}}{HE_{i,t}} \right)}{\left(\frac{Y_{i,t}}{H_{i,t}} \right)} \right) \quad (9)$$

Waarbij LC staat voor de reële arbeidskosten en HE het totaal aantal gewerkte uren door werknemers. De totale arbeidskosten omvatten niet alleen de loonkosten, maar ook de indirecte belastingen en premies en subsidies. In de noemer staat de reële arbeidsproductiviteit per gewerkt uur, gemeten als bbp (Y) gedeeld door het totaal aantal gewerkte uren van alle werkenden (H).

Schattingresultaten

In tabel 1 zijn de schattingresultaten van onze drie modellen weergegeven. In de DOLS-schatting hebben alle onafhankelijke variabelen een statistisch significant effect op de kennisintensiteit van de sectorstructuur en laten het verwachte teken zien. Het intrinsieke effect en publieke R&D hebben beide een positief effect. Een verbetering van het intrinsieke effect met 1%-punt leidt tot een verbetering van het sectorstructureffect met 0,3%-punt. Een verbetering van de publieke R&D-intensiteit ten opzichte de benchmark verbetert het sectorstructureffect met 0,5%-punt.

Tabel 1: Schattingresultaten

Coëfficiënten en variabelen		Verklarende variabele: S_t		
		DOLS ^a	ECM ^b	GMM ^b
c	Constante	-	0,02 (0,64)	-
$I_{i,t-2}$	Intrinsiek effect: verschil t.o.v. benchmark	0,31** (3,37)	0,40** (3,14)	-
$PUB_{i,t-2}$	Publieke R&D: verschil in publieke-uitgaven als % van bbp t.o.v. benchmark	0,50** (2,33)	0,11 (0,17)	-
$O_{i,t-2}$	Openheid van de economie: verschil t.o.v. benchmark	0,03** (4,81)	0,10** (3,00)	-
$ULC_{i,t-2}$	Reële arbeidskosten per eenheid product: niveaus	-1,62** (-2,13)	-1,53 (-0,90)	-
λ	Foutcorrectieterm	-	-0,10** (-4,47)	-
$\Delta(S_{i,t-1})$	Autoregressieve term		0,17** (2,79)	0,47** (15,56)
$\Delta(I_{i,t-1})$	Mutatie van intrinsieke effect	-	0,32** (2,43)	0,15** (6,75)
$\Delta(PUB_{i,t-1})$	Mutatie van publieke R&D	-	0,12* (1,91)	0,57** (5,20)
$\Delta(O_{i,t-1})$	Mutatie van openheid	-	0,05** (3,87)	0,04** (5,64)
$\Delta(ULC_{i,t-1})$	Mutatie van arbeidskosten p.e.p.	-	-0,21 (-0,89)	-0,46 (-1,34)
Fixed effects landen?				
		-	JA	-.
Fixed effects tijd?				
		-	JA	-
R ²		0,99	0,44	
J-statistiek		-	-	13,8
Durbin-Watson statistiek		-	2,08	-
Periode		1991-2017	1992-2017	1991-2017

Bron: RaboResearch. Toelichting: publieke R&D is één jaar vertraagd. ^b Publieke R&D in niveaus is drie jaar vertraagd en in eerste verschillen vier jaar. ^{b, c} Openheid in eerste verschillen is niet vertraagd. ^c Behalve openheid zijn de andere onafhankelijke variabelen twee jaar vertraagd. Vertragingen zijn gebaseerd op economische intuïtie en optimale fit.

Het error-correctiemodel laat in niveaus vergelijkbare coëfficiënten zien als de DOLS-schatting, met uitzondering van publieke R&D. Bovendien zijn de effecten van publieke R&D en arbeidskosten statistisch niet-significant. De foutcorrectieterm (λ) vertoont een statistisch significant negatief teken, wat betekent dat kortetermijnafwijking van variabelen van het langetermijnevenwicht teniet gedaan worden in plaats van verder versterkt. De Durbin-Watson statistiek geeft aan dat er geen sprake is van autocorrelatie in de residuen in ons ECM model, terwijl de J-statistiek bij ons GMM model een indicatie is dat er geen sprake is van een overschot aan restricties. Ook in dynamische sfeer zijn de effecten van alle onafhankelijke variabelen, behalve de arbeidskosten per eenheid product, statistisch significant. Dat beeld wordt bevestigd door schattingen met het GMM-model.

Ter controle hebben we de in onze analyse ook menselijk kapitaal meegenomen als aparte verklarende variabele en als interactieterm met het intrinsieke effect en publieke R&D. Hoewel in onze schattingen deze variabelen wel statistisch significant zijn, is de impact op onze scenario analyse beperkt. Daarom geven wij de voorkeur aan een simpeler model en nemen menselijk kapitaal niet mee in het model dat we gebruiken voor onze scenarioanalyse.