WORKING PAPER 5-15

Evaluation of federal tax incentives for private R&D in Belgium: An update

June 2015

Michel Dumont, dm@plan.be

Avenue des Arts 47-49 – Kunstlaan 47-49
1000 Brussels

E-mail: contact@plan.be
http://www.plan.be

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency.

The FPB performs research on economic, social-economic and environmental policy issues. For that purpose, the FPB gathers and analyses data, examines plausible future scenarios, identifies alternatives, assesses the impact of policy measures and formulates proposals.

The government, the parliament, the social partners and national and international institutions appeal to the FPB's scientific expertise. The FPB provides a large diffusion of its activities. The community is informed on the results of its research activities, which contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic Enterprise label (three stars) for its environmental policy

url: http://www.plan.be
e-mail: contact@plan.be

Publications

Recurrent publications:
Outlooks

Planning Papers (last publication):
The aim of the Planning Papers is to diffuse the FPB's analysis and research activities.

114 Les charges administratives en Belgique pour l'année 2012 / Administratieve lasten in België voor het jaar 2012
Chantal Kegels - February 2014

Working Papers (last publication):

4-15 Potential output growth in Belgium since the crisis: Lower and more uncertain
Igor Lebrun - June 2015

Responsible publisher: Philippe Donnay
Legal Deposit: D/2015/7433/19

Federal Planning Bureau
Avenue des Arts - Kunstlaan 47-49, 1000 Brussels
phone: +32-2-5077311
fax: +32-2-5077373
e-mail: contact@plan.be
http://www.plan.be

Evaluation of federal tax incentives for private R&D in Belgium: An update

June 2015

Michel Dumont, dm@plan.be

Abstract - This paper presents the results of a second evaluation of the tax incentives that were introduced – between 2005 and 2008 – by the Belgian federal government to support R&D activities of private companies. Compared with the first assessment, carried out in 2012, this evaluation extends the period considered by two years (2010 and 2011) and provides the results of a first assessment of the tax credit for investment in R&D and the tax deduction of 80% of qualifying gross patent income that were introduced in 2007. The second evaluation also elaborates on the difficulties of estimation procedures to establish the "causal" effect of public support and the importance to account for the strong persistence in firm-level R&D expenditures.

Jel Classification - H32, O32, O38

Keywords - R&D, tax incentives, subsidies, impact assessment

Table of contents

Executive summary 1

Synthèse 4

Synthese 7

1. Introduction 10

2. Data 15

3. Input additionality 22

3.1. Baseline specification 24

3.1.1. Persistence of R&D activities and public support 29

3.1.2. Policy Mix 36

3.1.3. Alternative dependent variables 38

3.1.4. Industry and firm heterogeneity 41

3.2. Selection bias 49

3.3. Instrumental variable estimation 54

3.4. Matching 59

3.5. Dynamic panel 64

3.6. Starting R&D activities 68

3.7. Bang for the Buck (BFTB) 70

4. Behavioural and output additionality 71

4.1. Orientation of R&D activities 71

4.2. Composition of R&D personnel 73

4.3. Output of R&D activities 76

5. Conclusions 79

References 81

Annex 90

Annex 1: Technical details on estimation procedures 90

Annex 2: Descriptive statistics 101

Annex 3: Results by industry 103

List of tables

Table 1 Average and median amount of public support for R&D in 2011 ... 16
Table 2 Share of each quartile in R&D expenditures and public support for R&D (2011) ... 17
Table 3 Responses of firms with public support as to R&D expenditures in 2011 (2012 R&D Survey) ... 18
Table 4 Evolution in the share of R&D active firms that receive public support ... 18
Table 5 Policy mix: combinations of public support for R&D ... 19
Table 6 Descriptive statistics by support scheme (2011) - all amounts in 1000 euro ... 20
Table 7 Results of the baseline panel estimation (2003-2011) ... 27
Table 8 Results of balanced panel estimation (2003-2011) ... 29
Table 9 Persistence of R&D activities (2003-2011) ... 30
Table 10 Persistence of public support for R&D (2003-2011) ... 31
Table 11 Correlation over time of R&D expenditures and public support ... 33
Table 12 Results of alternative panel specifications with one lag for public support (2003-2011) ... 34
Table 13 Results of alternative panel specifications with two lags for public support (2003-2011) ... 34
Table 14 Results of alternative panel specifications with two lags (2003-2011) ... 35
Table 15 Results of the estimation of the Policy Mix of public support for R&D (2003-2011) ... 38
Table 16 Results of panel estimation with alternative dependent variables (2003-2011) ... 40
Table 17 Results of panel estimation by firm size (2003-2011) ... 43
Table 18 Results of panel estimation by firm age (2003-2011) ... 44
Table 19 Results of panel estimation by Pavitt category (2003-2011) ... 48
Table 20 Results of panel estimation by orientation of R&D activities (2003-2011) ... 49
Table 21 Determinants of receiving public support for R&D (2003-2011) ... 51
Table 22 Results of panel estimation accounting for (self-) selection (2003-2011) ... 53
Table 23 List of instruments ... 56
Table 24 Results of instrumental variable estimation (2003-2011) ... 57
Table 25 Average treatment effect on the treated – Nearest neighbour matching (2003-2011) ... 60
Table 26 Results of dynamic panel estimation (2003-2011) ... 65
Table 27 panel unit root tests of levels and first differences (2003-2011) ... 67
Table 28 Results of an Error Correction Model (2003-2011) ... 68
Table 29 Results of an estimation of the decision to start R&D activities (2003-2011) ... 69
Table 30 Bang for the Buck for significant elasticity estimates in alternative estimation procedures (2003-2011) ... 70
Table 31 Results of panel estimation by orientation of R&D activities (2003-2011) ... 72
Table 32 Results of panel estimation of the composition of R&D personnel (2003-2011) ... 75
Table 33 Results of a panel estimation of the effects of support for R&D on productivity (2003-2011) ... 77

List of figures

Graph 1 Evolution of Gross Expenditures on R&D as a percentage of GDP (2000-2012) 10
Graph 2 Growth in government support (2006-2011) and GERD (2006-2012) 12
Graph 3 Budgetary cost of partial exemption from advance payment of withholding tax 15
Graph 4 Budgetary cost of tax credit for R&D investment and tax deduction 80% patent income 16
Graph 5 Evolution of R&D expenditures (1992-2011) in the five industries with highest R&D expenditures (2011) 24
Graph 6 Impact by quantile: regional subsidies and partial exemption R&D cooperation 45
Graph 7 Impact by quantile: partial exemption YIC and researchers with a PhD or civil engineer degree 45
Graph 8 Impact by quantile: partial exemption researchers with a master degree and tax credit R&D 46
Graph 9 Impact by quantile: tax deduction of 80% of gross patent income 46
Graph 10 Dose response function regional subsidies 62
Graph 11 Dose response function partial exemption R&D cooperation 62
Graph 12 Dose response function Young Innovative Companies 62
Graph 13 Dose response function partial exemption researchers with a PhD or civil engineer degree 63
Graph 14 Dose response function partial exemption researchers with a Master degree 63
Graph 15 Dose response function Tax credit R&D investment 63
Graph 16 Dose response function Tax deduction 80% patent income 64
Graph 17 Evolution of R&D personnel by level of education (Business Sector in Belgium) 74

Executive summary

In view of its commitment to raise Research and Development (R&D) expenditures to 3 per cent of GDP, the Belgian federal government introduced – between 2005 and 2008 – a number of tax incentives in support of the R&D activities of private firms. Whereas R&D intensity in Belgium decreased substantially after 2001, it started rising again from 2005 onwards, reaching a record high of 2.24 per cent of GDP in 2012. The increase coincides with a strong uptake in the use by private firms of the tax benefits for R&D. Given the substantial budgetary cost of the tax benefits, the question remains to what extent the tax incentives have been instrumental in raising the R&D expenditures of private firms.

In 2012, a first evaluation of the tax incentives for R&D indicated that the partial exemption from advance payment of the withholding tax on the wages of researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineering degree and on the wages of R&D personnel of companies that cooperate in research with a university, higher education institution or scientific institution appeared to be most effective in stimulating firms to perform more R&D. Effectiveness of support was found to be somewhat lower for the partial exemption from advance payment of the withholding tax for so-called Young Innovative Companies and for researchers with a master degree. The results also suggested that the combination of different support schemes tends to reduce the effectiveness of public support.

The first evaluation covered the period 2001-2009. As the tax benefits were introduced between 2005 and 2008, information on tax support was only available for a couple of years. This paper presents the results of a second evaluation, based on a dataset that includes two additional years (2010 and 2011). In contrast with the first assessment, this evaluation also evaluates the effects of the two incentives that were introduced most recently, in effect, the tax credit for R&D investment (introduced in 2007) and the tax deduction of 80% of gross patent income (effective as of tax year 2008). Given the importance of regional R&D subsidies in Belgium, the total amount of direct support provided by regional funding agencies is controlled for in the assessment of the effects of tax support on private R&D activities.

This paper also elaborates more on the difficulties of estimation procedures to establish a causal link between public support for R&D and own R&D expenditures of private firms. Whether it concerns direct support (subsidies) or tax benefits, it is always the private company that decides autonomously how much it invests in R&D and whether or not to apply for a subsidy or a tax benefit. Moreover, public agencies follow explicit rules to grant subsidies. For example, most regional agencies in Belgium have programmes that specifically target SMEs or technology fields. Subsidies are granted, based on the quality of the project proposal which may reflect firm-specific characteristics known to the agency or the reviewers but not to the evaluator. From the perspective of the evaluator of public support the ideal would be to grant support fully randomly and after some time assess the difference in R&D expenditures and behaviour between the group of firms that received support and the group of firms that did not receive support. Rather than assisting to this "ideal experiment", the evaluator mostly needs to rely on information gathered after support has been given in a non-random way. Estimation of the impact of public support on private R&D expenditures therefore needs to take into account firm characteristics

that may affect the probability of firms to apply for and receive support. As most econometric procedures that try to tackle the problem of (self-) selection and endogeneity of public support have known limitations, this paper compares the results of a baseline panel specification with the results from alternative estimators and highlights robust findings.

The most robust results indicate that the partial exemption from advance payment of the withholding tax on the wages of researchers with a master degree and the direct support (subsidies) provided by regional funding agencies are most effective in raising private R&D expenditures. The statistically significant positive impact, in the baseline specification, of the partial exemption from advance payment of the withholding tax on the wages of researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineering degree and on the wages of R&D personnel of companies that cooperate in research with a university, higher education institution or scientific institution, is not confirmed by alternative estimates that account for selection and endogeneity.

There are almost no indications that the tax benefit (partial exemption) for Young Innovative Companies, the tax credit for R&D investment and the tax deduction of 80% of gross patent income have a statistically significant effect on private R&D expenditures. Estimates of the impact of public support on the orientation of R&D activities, however, suggest that the partial exemption for Young Innovative Companies, the tax deduction of 80% of gross patent income and regional subsidies induce firms to shift their R&D activities away from development towards research (basic or applied). This result seems relevant in view of the recent reports – to some extent corroborated in this paper – that large companies increasingly shy away from scientific research which they leave up to small start-ups.

In line with results in the first evaluation, estimates reported in this paper suggest that the combination of different support schemes reduces the effectiveness of public support. This appears to be the case for firms that combine direct support (subsidies) by the regions with a partial exemption from advance payment of the withholding tax for researchers with a master degree but also for several combinations of federal tax incentives. The reduction in effectiveness is substantial for firms that combine more than two support schemes.

The effectiveness of most support schemes does not appear to be constant over different rates of subsidization (ratio of support to R&D expenditures). There are some indications that the additionality of support decreases with increasing rates of support.

An assessment of the impact of public support on the educational mix of R&D personnel shows that the partial exemption from advance payment of the withholding tax on the wages of researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineering degree and the partial exemption from advance payment of the withholding tax on the wages of researchers with a master degree, to some extent, result in substitution of researchers with a high education degree (PhD or master) for researchers with a lower degree. For the partial exemption for researchers with a master degree the extent of the substitution even appears to result in an impact on the total number of R&D employees that is not statistically significant. The positive impact on R&D expenditures of this scheme may therefore be partly due to an increase in the average skill – and consequently wage – level of researchers although there are no indications of a statistically significant effect of any of the support schemes on the average wage of researchers.

For the partial exemption from advance payment of the withholding tax, as of 2014, preapproval by the Federal Public Planning Service Science Policy is required for new R&D projects and, from 2015, for existing projects. As the data considered in this paper cover the period 2003-2011, the effects of this preapproval could not be assessed and will be the subject of future evaluations.

Synthèse

Face à son engagement à porter les dépenses en recherche et développement (R&D) à trois pour cent du PIB, le gouvernement fédéral belge a introduit, entre 2005 et 2008, un certain nombre d'incitations fiscales afin de stimuler les activités de R&D dans les entreprises privées. Alors que l'intensité en R&D en Belgique a diminué sensiblement après 2001, elle a à nouveau augmenté à partir de 2005 pour culminer à 2,24 % du PIB en 2012. Cette augmentation coïncide avec une forte progression de l'utilisation des incitations fiscales à la R&D par les entreprises privées. Compte tenu du coût budgétaire important des incitations fiscales, la question qu'il convient de poser est dans quelle mesure elles ont permis d'accroître les dépenses de R&D des entreprises privées.

Une première évaluation des incitations fiscales à la R&D, réalisée en 2012, a montré que la dispense partielle de versement de précompte professionnel sur les salaires des travailleurs de la connaissance titulaires d'un doctorat en sciences (appliquées) ou en médecine (vétérinaire) ou d'un tire d'ingénieur civil et la dispense sur les salaires du personnel R&D d'entreprises qui collaborent dans le domaine de la recherche avec une université, un établissement d'enseignement supérieur ou une institution scientifique sont les mesures les plus efficaces pour encourager les entreprises à développer leurs activités de R&D. L'efficacité était légèrement inférieure pour la dispense partielle de précompte professionnel au bénéfice des jeunes entreprises innovantes et des travailleurs de la connaissance détenteurs d'un master. Les résultats ont aussi laissé apparaître que la combinaison de différents dispositifs de soutien tend à réduire leur efficacité.

La première évaluation a couvert la période 2001-2009. Au moment où cette première évaluation a été réalisée, les informations sur les incitations fiscales n'étaient disponibles que pour un nombre réduit d'années puisque les incitations ont été introduites entre 2005 et 2008. La présente étude expose les résultats d'une deuxième évaluation, basée sur une série de données qui inclut deux années supplémentaires (2010 et 2011). Contrairement à la première, la deuxième évaluation porte aussi sur les effets de deux incitations introduites plus récemment, soit le crédit d'impôt à l'investissement en R&D (instauré en 2007) et la déduction fiscale de 80 % des revenus bruts des brevets (en vigueur depuis l'année d'imposition 2008). Compte tenu de l'importance des subventions régionales à la R&D en Belgique, le montant total de l'aide directe octroyée par les organismes régionaux de financement est pris en compte dans l'estimation des effets des incitations fiscales sur les activités privées de R&D.

La présente étude aborde également plus en détail les difficultés d'établir, par le biais de méthodes d'estimation, un lien de cause à effet entre l'aide publique à la R&D et les dépenses de R&D des entreprises privées. Qu'il s'agisse d'une aide directe (subventions) ou d'incitations fiscales, l'entreprise privée décide toujours en toute autonomie du montant qu'elle investit dans la R&D et de l'opportunité de solliciter ou non les aides. En outre, les organismes publics octroient les subventions en s'appuyant sur des règles explicites. Par exemple, la plupart des agences régionales belges disposent de programmes qui ciblent spécifiquement les PME ou des domaines technologiques. Les subventions sont alors octroyées sur la base de la qualité de la proposition de projet, laquelle peut refléter des caractéristiques propres de l'entreprise, susceptibles d'être connues de l'agence régionale ou des rapporteurs des propositions mais pas de l'évaluateur de l'aide publique. Du point de vue de l'évaluateur, la situation idéale consisterait à octroyer l'aide de façon toute à fait aléatoire et, quelque temps après, d'évaluer les différences de comportement et de niveau de dépenses de R&D entre le groupe d'entreprises qui a reçu l'aide et le groupe qui ne l'a pas reçue. La plupart du temps, l'évaluateur ne peut prétendre à cette situation idéale et doit se fonder sur des informations recueillies après l'octroi non aléatoire de l'aide. L'estimation de l'impact de l'aide publique sur les dépenses privées de R&D doit dès lors prendre en compte les caractéristiques des entreprises qui sont susceptibles d'influer sur la probabilité pour une entreprise de solliciter une aide et de la recevoir. Dès lors que la plupart des exercices économétriques qui ont tenté d'éliminer le problème d'autosélection et d'endogénéité de l'aide publique ont rencontré des limites, la présente étude compare les résultats d'une spécification de référence avec les résultats d'autres estimations et expose uniquement les conclusions robustes à ce test.

Les résultats les plus robustes montrent que la dispense partielle de versement de précompte professionnel sur les salaires des travailleurs de la connaissance titulaires d'un master et l'aide directe (subventions) octroyée par les organismes régionaux de financement sont les mesures qui font le plus augmenter les dépenses privées de R&D. L'impact positif statistiquement significatif, observé dans la spécification de référence, de la dispense partielle de versement de précompte professionnel octroyée pour des travailleurs de la connaissance docteurs en sciences (appliquées) ou en médecine (vétérinaire) ou ingénieurs civils ou encore octroyée pour le personnel R&D d'entreprises qui collaborent dans le domaine de la recherche avec une université, un établissement d'enseignement supérieur ou une institution scientifique n'est pas confirmé par d'autres estimations qui prennent en compte la sélection et l'endogénéité.

Pratiquement aucune indication d'un effet statistiquement significatif sur les dépenses de R&D n'a été observée pour la dispense partielle en faveur des Jeunes Entreprises Innovantes, le crédit d'impôt à l'investissement en R&D et la déduction fiscale de 80 % des revenus bruts des brevets. Toutefois, les estimations de l'impact de l'aide publique sur l'orientation des activités de R&D suggèrent que la dispense partielle en faveur des Jeunes Entreprises Innovantes, la déduction fiscale de 80 % des revenus bruts des brevets et les subventions régionales encouragent les entreprises à réorienter leurs activités de R&D vers la recherche (fondamentale ou industrielle) au détriment du développement expérimental. Le résultat selon lequel les grandes entreprises se désengagent progressivement de la recherche au profit des jeunes pousses – et qui est ici corroboré dans une large mesure – semble pertinent et est d'ailleurs déjà constaté dans la littérature récente.

Conformément aux résultats de la première évaluation, les estimations présentées dans cette étude suggèrent que la combinaison de différents régimes de soutien réduit l'efficacité de l'aide publique. Ce résultat semble se vérifier pour les entreprises qui combinent l'aide directe régionale (les subventions) et la dispense partielle de précompte professionnel pour les chercheurs détenteurs d'un master, mais aussi pour plusieurs combinaisons d'incitations fiscales fédérales. Cette baisse d'efficacité est substantielle pour les entreprises qui combinent plus de deux systèmes d'aide.

En outre, l'efficacité de la plupart des systèmes d'aide publique n'est pas constante pour les différents taux de subventionnement (ratio d'aide par rapport aux dépenses de R&D). Certains éléments tendent à indiquer que l'additionnalité de l'aide diminue à mesure que le taux de subventionnement augmente.

Une évaluation de l'impact de l'aide publique sur le bagage éducatif du personnel R&D montre que la dispense partielle de précompte professionnel pour les chercheurs docteurs en sciences (appliquées) ou en médecine (vétérinaire) ou ingénieurs civils et la dispense partielle de précompte professionnel pour les chercheurs titulaires d'un master débouchent sur un remplacement partiel des chercheurs titulaires d'un diplôme moins élevé par des chercheurs hautement qualifiés (titulaires d'un doctorat ou d'un master). En ce qui concerne la dispense partielle au profit des chercheurs titulaires d'un master, l'importance de la substitution semble avoir un effet non statistiquement significatif sur l'effectif total de personnel R&D. L'impact positif de ce système sur les dépenses de R&D pourrait dès lors être en partie dû à une hausse du niveau moyen de qualification – et par conséquent de la rémunération – des chercheurs, même s'il n'existe aucune indication d'un effet statistiquement significatif des systèmes d'aide sur la rémunération moyenne des chercheurs.

Toute entreprise qui souhaite bénéficier d’une dispense partielle de versement de précompte professionnel doit notifier, respectivement depuis 2014 et 2015, chaque nouveau projet de R&D et projet existant au service public de programmation politique scientifique en vue d'obtenir son approbation préalable. Étant donné que l'étude couvre la période 2003-2011, les effets de cette procédure n'ont pu être évalués mais le seront ultérieurement.

Synthese

Om tegemoet te komen aan haar verbintenis om de uitgaven voor onderzoek en ontwikkeling (O&O) te verhogen tot 3% van het bbp heeft de Belgische federale regering – tussen 2005 en 2008 – een aantal fiscale maatregelen genomen om de O&O-activiteiten van particuliere ondernemingen te ondersteunen. De Belgische O&O-intensiteit daalde aanzienlijk na 2001, maar nam opnieuw toe vanaf 2005 en bereikte in 2012 zelfs een recordhoogte van 2,24 procent van het bbp. Die stijging valt samen met een sterk toegenomen gebruik van de fiscale voordelen voor O&O door particuliere ondernemingen. Gelet op de aanzienlijke budgettaire kosten die aan die fiscale voordelen verbonden zijn, blijft de vraag in welke mate de fiscale voordelen hebben geholpen om de O&O-uitgaven van particuliere ondernemingen te verhogen.

In 2012 gaf een eerste evaluatie van de fiscale voordelen voor O&O aan dat de gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor kenniswerkers met een diploma van doctor in de (toegepaste) wetenschappen of de (dier)geneeskunde of burgerlijk ingenieur en voor het O&O-personeel van ondernemingen die samenwerken met universiteiten, hogescholen en wetenschappelijke instellingen heel doeltreffend lijkt om ondernemingen aan te moedigen meer aan O&O te doen. De gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor de zogenoemde Jonge Innoverende Ondernemingen en voor kenniswerkers met een masterdiploma was iets minder doeltreffend. De resultaten gaven ook aan dat de combinatie van verschillende steunprogramma's de doeltreffendheid van de overheidssteun doet afnemen.

De eerste evaluatie had betrekking op de periode 2001-2009. Aangezien de fiscale voordelen tussen 2005 en 2008 werden ingevoerd, was er slechts voor een aantal jaren informatie over de fiscale steun beschikbaar. De voorliggende paper presenteert de resultaten van een tweede evaluatie, gebaseerd op een tijdreeks die twee bijkomende jaren omvat (2010 en 2011). In tegenstelling tot de eerste evaluatie, worden hier ook de effecten van de twee meest recente stimuli geëvalueerd, namelijk het belastingkrediet voor O&O-investeringen (in 2007) en de belastingaftrek van 80% voor bruto-inkomsten uit octrooien (in voege vanaf het belastingjaar 2008). Gelet op het belang van gewestelijke subsidies voor O&O en innovatie in België, werd bij de schatting van de impact van de fiscale steun op particuliere O&O-activiteiten rekening gehouden met het totale bedrag aan directe steun dat ondernemingen ontvingen van de gewestelijke financieringsagentschappen.

Deze paper gaat ook dieper in op de moeilijkheden die gepaard gaan met econometrische schattingen om een oorzakelijk verband vast te stellen tussen overheidssteun voor O&O en de eigen O&O-uitgaven van particuliere ondernemingen. Ongeacht of het gaat om directe steun (subsidies) of fiscale voordelen, het is altijd de particuliere onderneming die autonoom beslist hoeveel er in O&O wordt geïnvesteerd en of er gebruik wordt gemaakt van subsidies of fiscale voordelen. Bovendien volgen overheidsinstellingen expliciete regels bij het toekennen van subsidies. De meeste gewestelijke financieringsagentschappen in België hebben bijvoorbeeld programma's die specifiek gericht zijn op kmo's en technologische domeinen. Subsidies worden verleend op basis van de kwaliteit van het projectvoorstel dat ondernemings-specifieke kenmerken kan weerspiegelen die de instelling en de projectbeoordelaars kennen, maar de evaluator van de overheidssteun niet. Vanuit het perspectief van de evaluator, zou de overheidssteun idealiter volledig willekeurig moeten toegekend worden en zou na enige tijd het verschil in O&O-uitgaven en gedrag tussen de ondernemingen die steun kregen en de ondernemingen die geen steun kregen, geëvalueerd moeten worden. In plaats van deel te kunnen uitmaken van dit ideaal 'experiment', moet de evaluator vooral steunen op de informatie die werd verzameld nadat de steun niet-willekeurig werd verleend. De schatting van de impact van overheidssteun op O&O-uitgaven moet daarom rekening houden met ondernemings-specifieke kenmerken die van invloed kunnen zijn op de mogelijkheid dat een onderneming steun aanvraagt en krijgt. Gezien de beperktheid van de meeste econometrische procedures die het probleem van (zelf-)selectie en endogeniteit van de overheidssteun trachten op te lossen, vergelijkt deze paper de resultaten van een basisspecificatie met de resultaten van alternatieve schattingen en vestigt de aandacht op de robuuste bevindingen.

De meest robuuste resultaten geven aan dat de gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor kenniswerkers met een masterdiploma en de directe steun (subsidies) van gewestelijke financieringsagentschappen het doeltreffendst zijn om particuliere O&O-uitgaven te verhogen. De statistisch significante positieve impact in de basisspecificatie, van de gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor kenniswerkers met een doctoraat in de (toegepaste) wetenschappen of de (dier)geneeskunde of met een diploma van burgerlijk ingenieur en voor O&O-personeel van ondernemingen die samenwerken met universiteiten, hogescholen en wetenschappelijke instellingen, wordt niet bevestigd in alternatieve schattingen die rekening houden met selectie en endogeniteit.

Er zijn vrijwel geen indicaties dat het fiscaal voordeel (gedeeltelijke vrijstelling) voor Jonge Innoverende Ondernemingen, het belastingkrediet voor O&O-investeringen en de belastingaftrek van 80% voor bruto-inkomsten uit octrooien een statistisch significant effect hebben op particuliere O&O-uitgaven. Schattingen van de impact van overheidssteun op de oriëntatie van O&O-activiteiten laten echter uitschijnen dat de gedeeltelijke vrijstelling voor Jonge Innoverende Ondernemingen, de belastingaftrek van 80% voor bruto-inkomsten uit octrooien en gewestelijke subsidies ondernemingen aansporen om hun O&O-activiteiten te verschuiven van ontwikkeling naar (fundamenteel of industrieel) onderzoek. Dit resultaat lijkt relevant in het licht van de recente bevindingen dat grote ondernemingen steeds meer terugschrikken voor fundamenteel of industrieel onderzoek en dit overlaten aan kleine startende ondernemingen (dit wordt in bepaalde mate bevestigd in deze paper).

In overeenstemming met resultaten uit de eerste evaluatie, suggereren de schattingen in deze paper dat de combinatie van verschillende steunprogramma's de overheidssteun minder doeltreffend maakt. Dat lijkt het geval voor ondernemingen die directe gewestelijke steun (subsidies) combineren met een gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor kenniswerkers met een masterdiploma, maar ook voor verschillende combinaties van federale fiscale stimuli. De doeltreffendheid neemt aanzienlijk af voor ondernemingen die meer dan twee voordelen combineren.

De doeltreffendheid van de meeste steunmaatregelen lijkt niet constant te zijn over verschillende subsidiëringspercentages (verhouding van steun tot O&O-uitgaven). Er zijn aanwijzingen dat de additionaliteit van steun afneemt bij hogere steunpercentages.

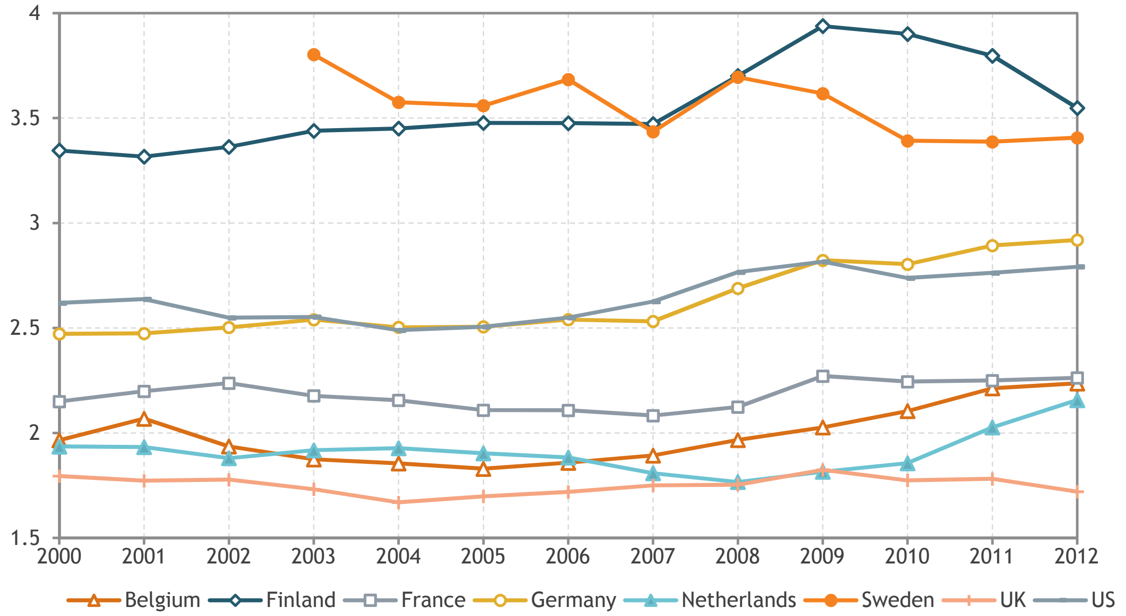
Een schatting van de impact van overheidssteun op de samenstelling van het O&O-personeel volgens scholingsniveau toont dat de gedeeltelijke vrijstelling van doorstorting van bedrijfsvoorheffing voor kenniswerkers met een doctoraat in de (toegepaste) wetenschappen of de (dier)geneeskunde of een diploma van burgerlijk ingenieur en de gedeeltelijke vrijstelling van bedrijfsvoorheffing voor vorsers met een masterdiploma, in zekere mate, resulteert in de vervanging van onderzoekers met een lager diploma door onderzoekers met een hoger onderwijsdiploma (doctor of master). Voor de gedeeltelijke vrijstelling voor vorsers met een masterdiploma lijkt de omvang van de vervanging zelfs te leiden tot een effect op het totaal aantal O&O-werknemers dat niet statistisch significant is. De positieve impact op de O&O-uitgaven in dit programma kan daarom deels toe te schrijven zijn aan de toename van het gemiddeld scholingsniveau – en bijgevolg loonniveau – van de onderzoekers, hoewel er geen indicaties zijn van een statistisch significant effect van een van de steunmaatregelen op het gemiddeld loon van de onderzoekers.

Voor de gedeeltelijke vrijstelling van bedrijfsvoorheffing is vanaf 2014 de voorafgaande goedkeuring van de Programmatorische Overheidsdienst Wetenschapsbeleid vereist voor nieuwe O&O-projecten en vanaf 2015 voor bestaande projecten. Aangezien de gegevens in deze paper de periode 2003-2011 bestrijken, konden de effecten van deze voorafgaande goedkeuring niet worden geëvalueerd en zullen ze het voorwerp uitmaken van toekomstige evaluaties.

1. Introduction

At the 2002 Summit of Barcelona, EU member states decided to set the target of how much a country should spend on Research and Development (R&D) – by 2010 – at 3 percent of GDP. Graph 1 shows the evolution of Gross Expenditures on R&D (GERD) in per cent of GDP, for seven EU countries and the US, over the period 2000-2012. As can be seen, only Finland and Sweden witnessed R&D intensity above the 3% target in 2010. The Europe 2020 strategy, launched in 2010, reasserted R&D intensity of 3 per cent as one of its five headline targets.[1] Having reached a peak in 2001, R&D intensity in Belgium dropped continuously until 2005, after which it increased steadily, even after the post-2008 economic slowdown which substantially affected the R&D intensity of other countries (for example, Finland, Sweden but also the US). In 2012, R&D expenditures in Belgium represented 2.24 percent of GDP, a record high. Of the eight countries considered in graph 1, Belgium witnessed the highest relative growth in GERD, after Germany. Over the period 2006-2012 Belgium's relative growth was higher than in any of the other countries considered.

Graph 1 Evolution of Gross Expenditures on R&D as a percentage of GDP (2000-2012)



Source: OECD Main Science and Technology Indicators

In view of its commitment to the 3% target, the Belgian federal government introduced a number of new tax incentives in support of R&D activities of private companies. In 2005 companies could apply for a partial exemption from advance payment of the withholding tax on the wages of R&D personnel involved in research cooperation with a university, a higher education institution or a scientific institution.

[1] Not all EU Member States adopted the EU-wide target of 3%. Finland and Sweden set a 4% target whereas the Netherlands set a target of 2.5%. Belgium adopted a 3% target for 2020.

In July 2006, so-called Young Innovative Companies (YIC)[2] became eligible for a partial exemption for their R&D personnel.[3] Two other measures were introduced to permit R&D companies to receive partial exemption from advance payment of the withholding on the wages of researchers with a specific degree; for researchers with a PhD in exact or applied sciences, PhD in (veterinary) medicine or a civil engineering degree (introduced in 2006) and for researchers with a master degree in sciences – except for social and human sciences – which was introduced in 2007.

Starting in tax year 2008, the federal government grants a deduction of 80% of qualifying gross patent income from the taxable basis. With a statutory corporate income tax rate of 33.99%, this implies effective taxation of 6.8% on patent income (for example from licensing to third parties).

As of tax year 2007, Belgian companies can choose between a tax deduction or a tax credit for investment in R&D (tangible and intangible fixed assets and patents). If companies cannot use the entire tax credit against taxable income, the part that has not been used within five years, is refunded.

OECD (2013) ranks Belgium first in a group of 23 OECD countries, as to the increase in tax incentives for R&D, with a reported annual growth rate between 2006 and 2011 of 51.3%. In a group of 36 countries, Belgium ranks fourth in terms of tax incentives for R&D as a percentage of GDP and a shared fourteenth position in terms of direct funding for R&D. With an annual growth rate of 5.7%, Belgium ranks 10th in terms of the increase in direct funding for R&D. The scatter plot in graph 2 shows the correlation between the annual growth rate in, respectively, direct funding and tax incentives for R&D between 2006 and 2011 and the relative growth in GERD as % of GDP between 2006 and 2012. The graph, which considers the 18 OECD countries for which all data are available in OECD (2013), reveals a positive correlation between government support and R&D intensity which is especially strong for direct funding. The graph also shows that Belgium is an outlier in terms of its impressive growth in tax incentives, with a position somewhat below the average relationship between growth in tax incentives and growth in GERD in per cent of GDP. Correlations do not provide any evidence as to causality, but graph 2 seems to suggest that direct funding is more effective in raising R&D expenditures. Both correlations are however strongly affected by the very strong relative growth in GERD of Slovenia (76.67% between 2006 and 2012), which over the period 2006-2011, ranks first in growth in direct funding (annual growth rate of 40.3%) but only 10th in terms of growth in tax incentives (annual growth rate of 7.3%).[4] For a group of 27 OECD countries, the correlation between direct funding for R&D as % of GDP and Business Expenditures on Research and Development (% of GDP) in 2011 is 0.49, which is statistically significant.

---

2 A Young Innovative Company is defined (see Belgian Science Policy, 2006) as a company which:
- carries out research projects;
- has been set up for less than 10 years before January 1 of the year during which the advance payment exemption is granted;
- is not set up within the framework of concentration, a restructuration, an extension of a pre-existing activity or resumption of such activities;
- has made expenditures on R&D representing at least 15% of the total costs in the foregoing taxable period.

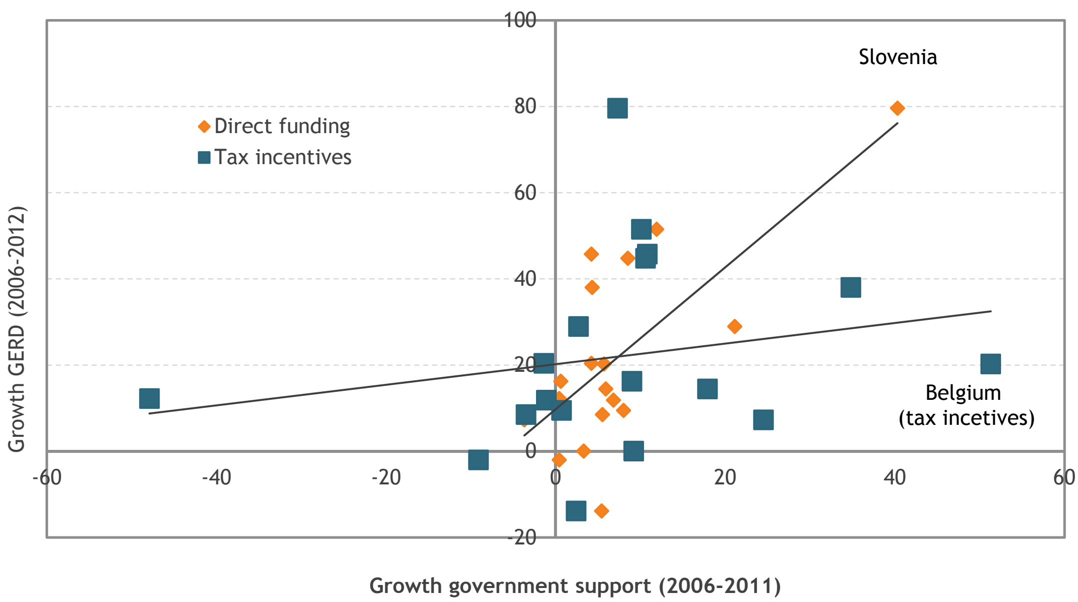
3 In 2013, the European Commission launched an investigation to evaluate whether the scheme of partial exemption for YIC was in line with the EU rules on state aid. One of the issues was that the legal basis for the scheme did not contain any reference to the eligible categories of research. In January 2015, the European Commission decided that the scheme complied with the old rules on state aid but expected that the legal basis is aligned with the new rules on state aid for Research, Development and Innovation that were adopted in 2014.

4 Leaving out Slovenia substantially reduces the correlation between direct funding and GERD and slightly increases the correlation between tax incentives and GERD although the latter correlation is still lower.

The correlation between tax incentives and BERD (both in in % of GDP) is also positive (0.22), but not statistically significant.

The top three countries[5] in terms of R&D intensity (BERD as % of GDP) actually have no tax incentives for R&D activities of private companies but only provide direct funding. Although there appears to be a shift from direct towards indirect public support, Busom et al. (2015) point out that there is yet no clear evidence as to which type of support is most effective in raising private R&D activities as most studies do not evaluate different instruments but tend to focus on a single form of support (subsidy or tax benefit).

Graph 2 Growth in government support (2006-2011) and GERD (2006-2012)



Source: OECD Science, Technology and Industry Scoreboard 2013

The most recently introduced tax incentive for R&D, the tax deduction of 80% of gross patent income fits in an international trend to provide tax benefits for intellectual property rights. In 2014, according to Evers et al. (2015), 12 European countries provided reduced corporate tax rates for income derived from intellectual property (IP). The schemes in Belgium, France and the UK only apply to patents, in contrast with other countries which consider a broader base of intellectual property. With a patent box rate of 6.8% Belgium is in the middle in terms of generosity, between the 0% rate in Malta and a rate of 16.76% in France. Evers et al. (2015) provide a comparison of the different existing IP (patent) box regimes in Europe.

The European Commission and the OECD are generally opposed to patent box incentives given the potential of harmful tax competition (Evers et al. 2015: p. 504). *The OECD/G20 Base Erosion and Profit Shifting Project* recently issued a proposal on a common approach for IP regimes that aims at providing safeguards against profit shifting and at ensuring equal treatment across all sectors and businesses of

[5] Finland, Japan and Sweden.

different sizes. The proposal also states that new rules should be implemented that are consistent with existing OECD rules on the phasing out of harmful regimes (OECD 2015).

Griffith et al. (2014) find that firms tend to locate their intellectual property in countries with a generous corporate tax regime for IP. As several countries have patent box regimes, this may result in tax competition. Simulations suggest that the introduction of a paten box in the Benelux countries, resulted in a large and statistically significant increase in the share of new patents located in Belgium and the Netherlands (based on legal ownership). The introduction of a patent box in the UK in 2013 resulted in a decrease of these shares although the shares remain higher than before the introduction of a patent box in the Benelux countries. The simulations, however, also clearly indicate that the introduction of a patent box results in a substantial loss in government revenues which outweigh the increase in patent income in many countries. Of the countries considered in the simulations, the loss in government revenue is estimated to be the highest in Belgium.

This paper provides the results of a second evaluation of the tax incentives introduced by the federal government in Belgium, between 2005 and 2008, in support of R&D activities by private companies. In contrast with the first evaluation (Dumont, 2013), the effects of the two most recent support schemes – the tax credit for R&D investment and the tax deduction of 80% of gross patent income – are also assessed. Given the longer period for which data on the tax incentives are available, this paper reports the results of a first assessment of the dynamic (long-term) effects of the tax benefits for R&D although given that the period under consideration is still rather short, these results should be interpreted with caution.

As of income year 2014, the partial exemption from advance payment of the withholding tax is conditional on preapproval by the Federal Public Planning Service Science Policy. For new R&D projects and programmes registration is required as of January 2014. For existing projects and programmes, firms had until December 2014 to register. The data considered in this paper cover the period 2003-2011. The effects of the preapproval could therefore not be assessed in this evaluation.[6]

The second evaluation delves into the difficulties of estimation procedures to establish the causal impact of public support, based on observational data. All methods that try to account for the potential bias of estimates due to the fact that firms decide autonomously to apply for support and that this support is not granted randomly are known to have limitations. The research strategy adopted in this paper is therefore to compare the results of a baseline panel specification to the results of alternative estimates and to highlight the results that appear to be robust. Technical details on estimation procedures are avoided as much as possible in the main text and provided in Annex 1.

Chapter 2 shows the recent evolution of R&D activities and public support for R&D in Belgium and provides a description of the dataset that is used to evaluate the impact of the tax benefits for R&D.

Chapter 3 discusses the estimation of input additionality, in effect, the extent to which public support stimulates private firms to spend more on R&D than they would without support. The chapter reports

---

6 See for more details (e.g. the difference between a R&D project and a R&D programme): http://www.belspo.be/belspo/organisation/fisc_en.stm

the results of the baseline panel specification which are compared with the results of alternative estimation procedures and some robustness tests. The chapter ends with estimates of the so-called Bang for the Buck (BFTB), which provides an indication of how much additional own R&D expenditures results from 1 euro received in support.

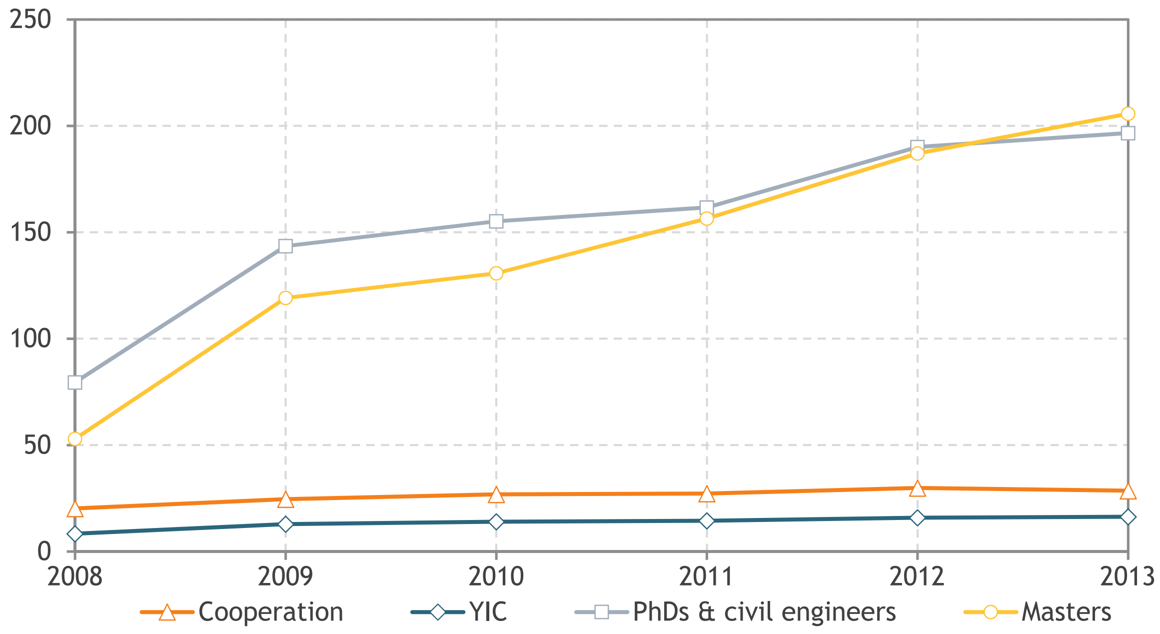
Chapter 4 considers the potential impact of public support on behavioural aspects of R&D activities, such as how much of their R&D firms spend on basic or applied research and the educational mix of the R&D personnel; and the impact on productivity. Conclusions are provided in chapter 5.

2. Data

The data used in this evaluation are retrieved from the *Policy Mix R&D* database, which was created by the Federal Public Service Finance. In the database, information from the Belgian biennial R&D survey, carried out by the Federal Science Policy Office, is linked to data on the direct support by the regions and data on the tax incentives provided by the federal government (partial exemption from advance payment of the withholding tax for R&D personnel; tax credit for investment in R&D and the deduction of 80% of qualifying gross patent income). The database also contains firm-level information from Bel-first, provided by Bureau van Dijk, based on annual accounts and balance sheets (for example, value added; cash flow; number of employees). The data in the first evaluation covered the period 2001-2009. For this evaluation, the period is extended with information for 2010 and 2011, from the R&D survey of 2012.

The different schemes of federal tax incentives for R&D by private companies have become increasingly popular, as can be seen in graphs 3 and 4, which show the total budgetary cost of the different support schemes.

Graph 3 Budgetary cost of partial exemption from advance payment of withholding tax
2008-2013

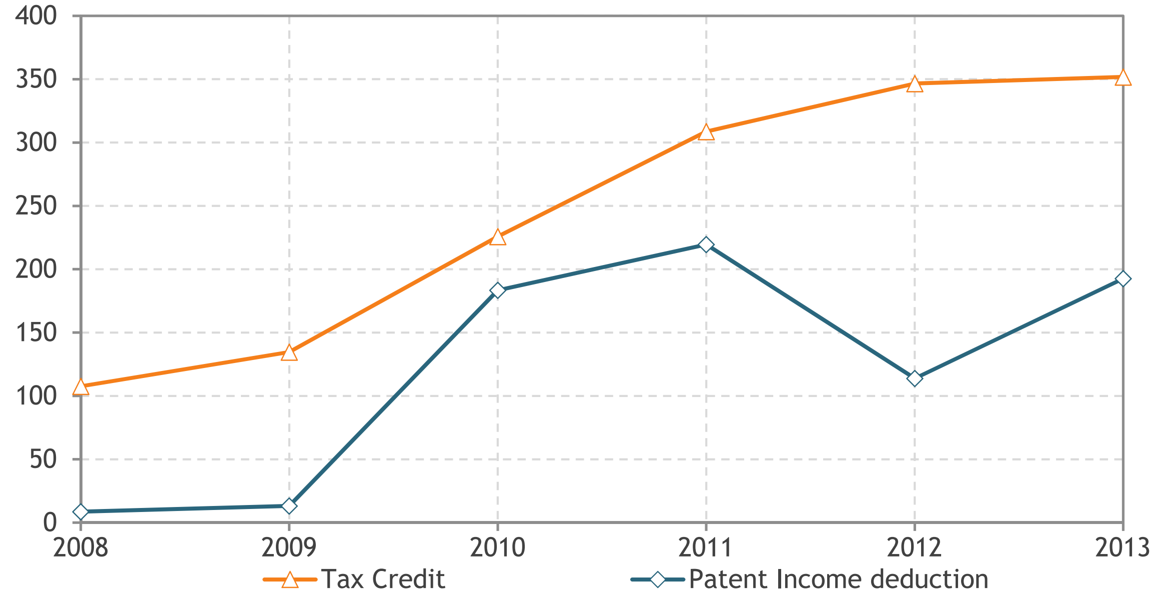


Source: Federal Public Service Finance

The budgetary cost of the four schemes of partial exemption from advance payment of the withholding tax on the wages of researchers increased from 161 million euro in 2008 to 447 million euro in 2013. In 2011, the cost of the partial exemption for researchers with a master degree caught up and in 2013 exceeded the cost of the partial exemption for researchers with a PhD or a civils engineering degree. As shown in graph 4, the budgetary cost of the tax credit of R&D investment and the tax deduction for 80%

of patent income increased even more considerably, from 147 million in 2008 to 544 million euro in 2012.[7]

Graph 4 Budgetary cost of tax credit for R&D investment and tax deduction 80% patent income
2008-2012



Source: Federal Public Service Finance

Table 1 shows the average and median amount of support received by companies in 2011 as well as the total number of firms that benefited from the given support scheme. In 2011, the largest number of firms received a subsidy, for an average of 296,023 euro. The partial exemption from advance payment of the withholding tax for researchers with a master degree has become more popular, in terms of total number of firms, than the partial exemption for researchers with a PhD or civil engineer degree although the average and median support of the latter is higher.

Table 1 Average and median amount of public support for R&D in 2011
In euro

	Average	Median	Number of firms
Regional subsidy	269,023	91,924	1,046
Research cooperation	71,996	15,866	214
Young Innovative Company	48,333	20,626	275
PhDs and civil engineers	205,463	38,655	689
Master	141,820	31,767	881
Tax credit R&D	1,940,991	64,487	159
Tax deduction 80% patent income	3,708,029	32,073	212

Note: The second column shows the average amount of public support for companies that benefited from the given instrument in 2011. The third column shows the median amount of public support and the final column the number of firms that benefited from the support.

[7] The strong dip, in 2011, in the budgetary cost of the tax deduction of 80% of gross patent income is due to an exceptional drop in patent income of a single company.

The tax credit for investment in R&D and the tax deduction of 80% of patent income are used by the smallest number of firms but the average support by far exceeds the average support of the other tax benefits and subsidies.

The substantial difference between the average and the median amount of support reveals the skewed distribution of public support, in line with the skewed distribution of R&D expenditures, shown in Table 2. The fourth quartile, grouping the 25% of firms with the highest R&D expenditures, in 2011 account for 94% of total R&D expenditures of private companies in Belgium and the third quartile for another 4%. The bottom half of firms performing R&D have a combined share of only 1% in total R&D expenditures. Table 2 also shows the distribution of public support for R&D in 2011 across quartiles. The scheme for Young Innovative Companies is the least skewed towards the highest quartile although the fourth quartile still accounts for 68% of the total amount saved by Young Innovative Companies through the partial exemption from advance payment of the withholding tax. Of all schemes of public support, regional subsidies most favour the first quartile, clearly reflecting the explicit targeting of SMEs by regional agencies. The tax credit for R&D investment and even more the tax deduction of 80% of gross patent income are extremely skewed towards the fourth quartile. For the tax deduction of patent income, the 25% of firms with the highest support account for 97% of total support. The skewness for the tax credit for investment in R&D and the tax deduction of 80% of patent income are also reflected in the large difference between the average and the median amount of support, reported in table 1. The last row in table 2 shows the share in total cash flow for the quartiles of R&D expenditures. Whereas the fourth quartile in terms of R&D expenditures accounts for 94% of total R&D expenditures, it only accounts for 85% of total cash flow.

Table 2 Share of each quartile in R&D expenditures and public support for R&D (2011)

	First quartile	Second quartile	Third Quartile	Fourth Quartile
R&D expenditures	0.00	0.01	0.04	0.94
Regional subsidy	0.04	0.06	0.11	0.80
Research cooperation	0.01	0.04	0.10	0.86
Young Innovative Company	0.02	0.08	0.22	0.68
PhDs and civil engineers	0.01	0.03	0.08	0.88
Master	0.01	0.04	0.11	0.84
Tax credit R&D	0.00	0.00	0.04	0.95
Tax deduction 80% patent income	0.00	0.01	0.03	0.97
Cash flow by quartile R&D expenditures	0.02	0.04	0.09	0.85

Note: The second up to the fifth column show the share of the first up to the fourth quartile in total R&D expenditures or the total amount of subsidies or tax benefits received by firms in 2011. The last row shows the share in total value added for each quartile of R&D expenditures.

Table 3 shows a breakdown of firms that benefited from public support in 2011, by response to the question in the 2012 R&D survey whether they performed R&D or not, in 2011. The fourth column shows the number of firms that received support but did not respond to the survey and the last column shows the number of firms that received support but that are not in the list of R&D firms, used by the Belgian Office of Science Policy for the biennial R&D survey. The relatively large number of firms that received a subsidy in 2011 but reported not to have performed R&D activities or are not listed in the list of R&D firms can be explained by the fact that these subsidies are not necessarily conditional upon R&D

activities as they more broadly aim to support innovation, with special attention to support SMEs without formal R&D activities. The fact that a substantial number of firms that received partial exemption from advance payment of the withholding tax for researchers in 2011 reported that they had no R&D activities in that year is more problematic as for these benefits R&D activities are a necessary condition for support.

Table 3 Responses of firms with public support as to R&D expenditures in 2011 (2012 R&D Survey)

	Performed R&D	Did not perform R&D	No response	Not in list R&D firms
Regional subsidy	412 (0.60)	62 (0.02)	382 (0.35)	95 (0.02)
Research cooperation	111 (0.65)	11 (0.02)	73 (0.24)	19 (0.09)
Young Innovative Company	96 (0.51)	17 (0.04)	124 (0.41)	38 (0.04)
PhDs and civil engineers	326 (0.69)	53 (0.02)	278 (0.28)	32 (0.01)
Master	379 (0.58)	83 (0.06)	353 (0.35)	66 (0.02)
Tax credit R&D	82 (0.82)	3 (0.03)	52 (0.12)	22 (0.03)
Tax deduction 80% patent income	69 (0.89)	2 (0.00)	48 (0.06)	99 (0.05)

Note: The table shows the response, of firms that received public support for R&D in 2011, on the question whether they performed R&D in 2011 (second column) or not (third column). The fourth column shows the number of firms that received support but did not respond to the survey and the final column shows the number of firms that are not included in the list of R&D firms. The numbers in brackets denote the share of each of the four groups in the total amount of support for the specific scheme, in 2011 (for example, the 412 firms that received a subsidy in 2011 account for 60% of the total amount of subsidies received by firms in 2011).

More than half of the companies that received a tax deduction of 80% of gross patent income, in 2011, are not listed in the list of R&D firms and another 5% explicitly report not have any R&D activities in Belgium. To benefit from this tax scheme, domestic companies or Belgian branches of foreign companies need to have a research centre in Belgium or abroad.[8] The numbers in table 3 suggest that the majority of firms benefiting from this measure do not perform research in Belgium, although these firms only account for 5% of the total amount saved by firms through the tax deduction for patent income.

Table 4 shows the trend in the share of firms doing R&D (according to their response in the biennial R&D survey) that receive some kind of public support for their R&D activities. Except for the partial exemption from advance payment of the withholding tax on the wages of researchers involved in research collaboration, the share of R&D firms that receive a tax benefit increased gradually between 2007 and 2011.

Table 4 Evolution in the share of R&D active firms that receive public support

	2007	2009	2011
Regional subsidy	0.25	0.18	0.26
Research cooperation	0.08	0.06	0.07
Young Innovative Company	0.04	0.05	0.06
PhDs and civil engineers	0.16	0.21	0.20
Master	0.07	0.17	0.24
Tax credit R&D	0.02	0.04	0.05
Tax deduction 80% patent income	-	0.02	0.04

Note: The table shows the share of firms that reported R&D activities in a given year that received a subsidy or a tax benefit in the same year.

8 As of 2013 SMEs are not required to have a research centre to benefit from the tax deduction of 80% of patent income.

In line with the evolution of the budgetary cost, shown in graph 3, the share of firms that receive partial exemption for researchers with a master degree, increased most, by 2011 surpassing the share of firms that benefit from partial exemption for researchers with a PhD or a civil engineer degree and almost equal to the share of firms that receive a subsidy. The share of R&D firms that receive partial exemption for researchers with a PhD or a civil engineer degree slightly decreased to 20% in 2011 from 21% in 2009. The increased use of the exemption for researchers with a master degree could be explained by the fact that as all researchers with a PhD or a civil engineer, by definition, also have a master degree, and as there is no difference in the rate of exemption between the two schemes, the exemption for researchers with a PhD or a civil engineer degree has become redundant. After a substantial drop in 2009, the share of firms receiving a subsidy from one of the regional agencies again equalled the 2007 level in 2011. Despite the relative increase, the share of R&D firms that receive a tax credit for R&D investment or a tax deduction of 80% of patent income remains low.

Table 5 provides an indication of the extent to which firms combine different forms of public support for R&D activities.

Table 5 Policy mix: combinations of public support for R&D

	Regional subsidy	Research cooperation	YIC	PhD	Master	Tax credit R&D	Tax deduction 80% patent income
Single use	0.54	0.38	0.54	0.38	0.39	0.23	0.34
Subsidy		0.20	0.26	0.08	0.08	0.06	0.07
Research cooperation	0.05		0.01	0.03	0.01	0.01	0.00
Young Innovative Company	0.08	0.01		0.01	0.00	0.10	0.02
PhDs and civil engineers	0.06	0.08	0.02		0.27	0.08	0.06
Master	0.06	0.03	0.00	0.26		0.05	0.08
Tax credit R&D	0.01	0.01	0.04	0.01	0.01		0.00
Tax deduction 80% patent income	0.00	0.00	0.00	0.01	0.01	0.00	
More than two	0.19	0.29	0.12	0.23	0.23	0.47	0.42

Note: The table shows the share of firms that receive, in a given year, only one of the given forms of public support (single use), combine it with one of the other benefits (second up to seventh line) or combine it with at least two other benefits (last line).

Table 6 provides descriptive statistics, for 2011, of the main variables in the dataset for those firms that responded to the R&D survey. The table shows the mean and the median of own R&D expenditures (net of public support); R&D intensity (R&D expenditures/ value added); the total amount of public support (subsidies and tax benefits) received by companies; the total full-time equivalent number of employees; firm age; cash flow and capital intensity (Tangible assets per employee).

Table 6 Descriptive statistics by support scheme (2011) - all amounts in 1000 euro

	No support		Subsidy		Exemption R&D cooperation		Exemption YIC	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
R&D expenditures	106	0	3,027	53	3,782	178	485	113
R&D/value added	0.00	0.00	0.02	0.00	0.02	0.00	0.01	0.00
Total public support	0	0	766	126	1.089	89	166	73
# employees (FTE)	70	20	169	22	208	46	18	7
Firm age	24	21	21	17	26	21	7	7
Cash flow	1,679	228	10,700	311	10,300	723	-34	61
Capital/employee	58	21	5	24	58	21	36	8
	Exemption PhDs and civil engineers		Exemption Master degree		Tax credit R&D investment		Tax deduction 80% patent income	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
R&D expenditures	5,850	535	3,702	334	13,700	1,145	15,700	329
R&D/value added	0.03	0.01	0.02	0.00	0.05	0.01	0.03	0.00
Total public support	1,094	134	715	95	2,988	360	3,261	23
# employees (FTE)	259	87	212	82	531	47	389	72
Firm age	26	22	26	22	25	11	27	22
Cash flow	20,500	1,280	13,400	1,259	49,800	66	49,400	2,161
Capital/employee	71	26	65	25	4	19	72	41

The statistics are shown for firms that did not receive any public support for R&D in 2011, firms that received a subsidy and firms that received one of the six tax benefits considered in this evaluation. The substantial difference between the average and the median reveals that the skewness of public support, as shown in table 1, parallels the skewness of most firm characteristics. Most variables are skewed to the right, with the mean exceeding the median. This reflects the fact that output and R&D is concentrated in a small group of large firms in most industries. The concentration is even more pronounced in R&D than in cash flow or employment. A notable exception is the cash flow of Young Innovative Companies. The average cash flow of firms that received support in 2011 under this scheme is actually negative whereas the median is slightly positive. Young Innovative Companies – by definition – spend much on R&D but often do not have much sales or even no sales at all. The negative average cash flow reflects a group of Young Innovative Companies known for their high cash burn rate, namely companies active in *Research and experimental development on biotechnology* (NACE 7211). From Table 6 it is clear that firms that do receive public support are, on average, larger and older than firms that do not receive support except for firms that receive a subsidy which tend to be larger but younger than firms that do not receive support and Young Innovative Companies, which – not surprisingly – are younger and smaller than non-supported firms or firms receiving other benefits.

Firms that receive a tax credit for R&D investment or a tax deduction of 80% of gross patent income are substantially larger, in terms of employment and cash flow but even more so in terms of R&D expenditures than firms without support. The statistics in table 6 show the average and median across industries.

The differences between the group of firms without support and the seven schemes of support may reflect the different extent to which industries are inclined to apply for support. Table A.2.1 in Annex 2 shows a table of standardized descriptive statistics. These have been computed by standardizing each variable with respect to the industry average (NACE two-digit).[9] As such the table shows the characteristics of firms without support, or firms that receive a subsidy or tax benefit, relative to the industry average. Overall, the statistics confirm the conclusions of table 6 that firms that receive public support tend to be larger than firms in the same industry that do not receive support.

The matching of the different data sources (R&D survey; subsidies and tax benefits, annual account and balance sheet data), permits to construct a panel for the period 2003-2011.[10] The time dimension of the panel is relatively short and is clearly dominated by the cross-section dimension (number of firms). Moreover the panel is not balanced, as there are several missing values for some years for a majority of firms. As shown in table A2.2, for only 3.88% of the firms, data are available for every year over the 2003-2011 period. The most frequent patterns of data availability is of firms with at most data for four years. Most panel estimation procedures allow for unbalanced panels, which implies that firms do not need to be dropped if they do not have data for all years although firms with few observations may not be very informative and it is useful to test the robustness of results for the whole sample by looking at the results of estimations for a panel with firms that have a sufficient number of observations over time.[11]

9 $Standardized\ X = \frac{(x - industry\ average(x))}{industry\ standard\ deviation\ (X)}$

10 Belfirst data, at most, go back 10 years in time from the most recent year so that earlier years can no longer be considered. The period considered predates the introduction of the federal tax incentives by two up to five years.

11 Estimation based on an unbalanced panel may suffer from panel attrition bias (see for example, Cameron and Trivedi (2010) and Cheng and Trivedi (2014)).

3. Input additionality

The evaluation of the impact of public support on the R&D activities of private companies is hampered by the limitations of econometric methods to establish causal links. Regression provides indications of association (correlation) between variables but does permit to prove indisputably any causal link. It is necessary to realize that a private company decides autonomously how much it invests in R&D. Availability of direct support or tax benefits is only one factor that companies take into consideration (see Becker (2013) for a recent survey on the determinants of R&D investment by private companies). It is also the company that decides to apply for a subsidy or a tax benefit. This complicates the assessment of the effects of introducing a support scheme or changing the conditions of public support. Moreover, public agencies follow explicit rules to grant subsidies. For example, most regional agencies in Belgium have programmes that specifically target SMEs or technology fields. Subsidies are granted, based on the quality of the project proposal which may reflect firm-specific characteristics known to the agency or the reviewers but most of the times not to the evaluator. More implicitly, agencies may favour a strategy of "picking the winner". The granting of subsidies or tax benefits is clearly subject to selection by agencies and self-selection by companies which implies that companies that receive public support for their R&D activities cannot be considered as a random sample of the population of companies (e.g. Lichtenberg 1984; Busom 2000; David et al. 2000; Klette et al. 2000; Wallsten 2000; Jaffe 2002; Cerulli 2010; Cantner and Kösters 2012).[12] If the selection by agencies that provide direct support and the self-selection and autonomy of firms to decide how much to invest in R&D are not accounted for, regression may result in a biased (optimistic) estimate of the causal impact of public support. Different estimation procedures exist to address selection bias and endogeneity of public support. All procedures have known advantages as well as several limitations. Unfortunately, no single method that is based on observational data can be considered to provide undisputed evidence on the causal effect of public support. The estimation strategy adopted in this paper is to start from a baseline panel estimation and to compare these results with other estimates to assess whether robust conclusions can be obtained (for the rationale of this strategy, see for example, Zúñiga-Vicente et al. 2014).

In this chapter, the intuition of the different estimation procedures is discussed with as little technical detail as possible. The interested reader is referred to Annex 1 for a more technical discussion of the estimations procedures used in this paper and an elaboration of the problems of econometric methods to establish causal links in the absence of the ideal fully randomized experiment.

Unbiased estimates of the impact of a given variable (in this paper, subsidies or tax benefits for R&D) on an outcome variable (private R&D expenditures) implies that no variable that may be relevant for the outcome variable, is omitted from the estimation. In a recent survey, Becker (2013) lists the main determinants of R&D expenditures that have been found in the literature:

– Individual firm or industry characteristics (internal finance and sales)

– Competition in product markets

[12] Antonelli and Crespi (2013) distinguish between vicious Matthew effects and virtuous Matthew effects of R&D subsidies. They present evidence for Italy that suggests the picking-the-winner strategy adopted by authorities positively contributed to the effectiveness of the subsidies.

- R&D tax benefits and subsidies
- Location and resource-related factors (e.g. spillovers from university research and research cooperation)
- Human capital embodied in knowledge workers
- Spillovers from foreign R&D

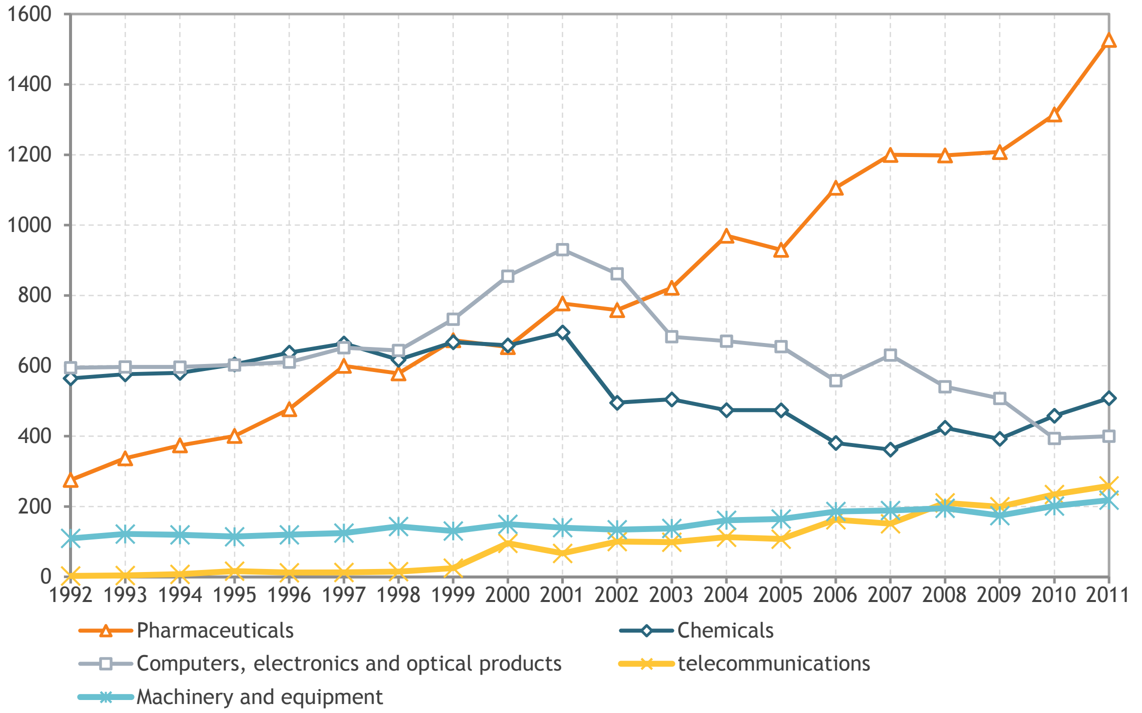
Cash flow is often used to proxy for credit constraints which are considered to be a major barrier to investment in R&D. Becker (2013) concludes that cash flow and sales are generally found to have a positive effect on R&D investment. This effect seems to be more important for small and young firms and for firms in the US and UK compared with firms in continental Europe. Brown et al. (2009, 2012) provide evidence, for the US and some EU countries, of the important role of cash flow but also of external equity in financing R&D. In a review of the literature on financing of R&D and innovation, Hall and Lerner (2010) conclude that there is rather strong evidence that small firms and start-ups in R&D-intensive industries face higher capital costs and are more dependent on venture capitalists than larger competitors, which seem to prefer internally generated funds to finance their investment. Kerr and Nanda (2014) point out that large firms can rely more on bank finance to finance innovation than small (young) firms as they can pledge tangible and intangible capital (e.g. patents) as collateral. Cincera et al. (2014) find that financing constraints affect young innovative companies more than older companies. Compared to US firms, R&D investment by young EU innovators appears to be more sensitive to cash flow, especially in medium- and high-tech industries.

Given data availability and the determinants put forward in the literature, four control variables are included in the estimation: cash flow, number of employees; firm age and capital intensity. Firm characteristics that do not change much over time, or at least over the period considered, can be reflected by firm fixed effects. In addition, region; industry (NACE two-digit) and year dummies can be included to account for differences across regions and industries and possible business cycle effects or other year-specific effects.

As graph 5 shows, there are substantial differences in the level but also in the evolution of R&D expenditures across industries, which reflect differences in terms of competition, technological opportunity, industry business cycles, credit constraints; adjustment costs and other industry-specific characteristics (see for example, David et al. 2000; Barlevy 2007; Ouyang 2011; Aghion et al. 2012; Cerulli and Poti 2014; Einiö 2014; Arora et al. 2015; Castellacci and Lei 2015; Thakor and Lo 2015). If these characteristics do not change over time, they can be captured by industry dummies.[13] Time-variant industry characteristics could be considered by using proxies such as output or a competition indicator (Herfindahl, concentration). The problem of the latter approach is that by using proxies the risk of biased estimates due to measurement error increases.[14] Moreover, it is not very likely that the proxies that are included will reflect all possible time-variant industry characteristics and the specification may therefore also suffer from an omitted variable bias due to unobserved industry heterogeneity. An alternative to control for time-varying industry-specific characteristics, used in this paper, is the inclusion of industry x year dummies, following the example of Aghion et al. (2012) and Einiö (2014).

13 David et al. (2000: p. 510) point out that industry-specific innovation opportunities are likely to change over time.
14 Becker (2013) points out that results appear to be sensitive to which measure of competition is considered.

Graph 5 Evolution of R&D expenditures (1992-2011) in the five industries with highest R&D expenditures (2011)



Source: OECD STAN R&D expenditures in Industry (ISIC Rev. 4), million euro (constant prices)

3.1. Baseline specification

The main research question in this paper concerns the input additionality of public support:

How much more do companies that receive a subsidy or tax benefit spend on their R&D activities than they would spend if they do not receive support?

To assess the impact of public support on the R&D expenditures of companies, panel regression is used as the baseline estimation as it permits to consider all available information on the amount of the different forms of support and R&D expenditures as well as the time dimension of public support (for example, the introduction of specific schemes and changes in the rate of partial exemption from advance payment of the withholding tax).

The baseline approach in this paper is what David et al. (2000) label as the "typical econometric approach", namely to regress some measure of private R&D on government funding, along with some other ''control'' variables. The estimates of a linear specification would provide a direct estimate of the Bang for the Buck, in effect, how much one euro in public support gives rise to how much additional own R&D expenditures by companies. However, as table 2 clearly shows, the distribution of R&D expenditures, as well as the distribution of public support for R&D, is extremely skewed and far from normal. The logarithm of R&D expenditures is relatively normally distributed (see graph A.1 in Annex 2).

Given the assumption, explicit in most estimation procedures, that errors are normally distributed, a log-linear specification is preferred (see also Aerts and Czaritzki 2004; Clausen 2008):

$$\begin{aligned}\ln\left(RD_{it}\right) = \alpha_0 + \beta^{reg}\ln\left(X_{it}^{reg}\right) + \beta^{coop}\ln\left(X_{it}^{coop}\right) + \beta^{YIC}\ln\left(X_{it}^{YIC}\right) + \beta^{PhD}\ln\left(X_{it}^{PhD}\right)\\ + \beta^{Master}\ln\left(X_{it}^{Master}\right) + \beta^{Credit}\ln\left(X_{it}^{Credit}\right) + \beta^{Patent}\ln\left(X_{it}^{Patent}\right) + \beta^{CF}\ln\left(CF_{it}\right)\\ + \beta^{E}\ln\left(Employees_{it}\right) + \beta^{A}Age + \beta^{KL}\ln\left(K_{it}/L_{it}\right) + \varepsilon_{it}\end{aligned} \quad (1)$$

Dependent variable:

RD_{it} : Internal R&D expenditures (excluding the amount of public support) of company *i* in year *t*

Explanatory variables (public support for R&D):

X_{it}^{reg} : Total amount of regional subsidies received company i in year t

X_{it}^{coop} : Total amount saved through partial exemption of the withholding tax on the wages of researchers cooperating with a university, college or a scientific institution

X_{it}^{YIC} : Total amount saved through partial exemption of the withholding tax on the wages of R&D personnel in Young Innovative Companies (YIC)

X_{it}^{PhD} : Total amount saved through partial exemption of the withholding tax on the wages of researchers with a PhD degree in exact or applied sciences, doctor degree in (veterinary) medicine or a civil engineering degree)

X_{it}^{Master} : Total amount saved through partial exemption of the withholding tax on the wages of researchers with a master degree (exception of master in social or human sciences)

X_{it}^{Credit} : Total amount saved through the tax credit for R&D investment

X_{it}^{Patent}: Total amount saved through the tax deduction of 80% of qualifying gross patent income

Control variables:

CF: Cash flow

Employees: Number of full time equivalent (FTE) employees

Age: Number of years since the company was incorporated

K/L: capital intensity (Tangible fixed assets per employee)

ε_{it}: error term (assumed to be randomly distributed with an expected value of 0 and a constant variance)[15]

The dependent variable in the baseline specification is total R&D expenditures reported by a company minus the total amount of public support for R&D received by the company. As pointed out by Cerulli (2010), if the amount of public support is known, 'own R&D' (total R&D minus public support) should be the target variable for the estimation of the input additionality of public support. If 'total R&D' (R&D expenditures including support) is used – which is the case in many empirical studies as data on the amount of support is not available – the fact that support is included in the target variable is a confounding element in assessing the effectiveness of public support (Cerulli, 2010: p. 427, see also among others David et al. 2000; Clausen 2008; Zúñiga-Vicente et al. 2014).

Most evaluations of public support for R&D consider a single instrument (subsidy or tax benefit). Busom et al. (2015) argue that as in most countries both direct and indirect support is provided, estimates that do not account for this policy mix could be biased. In this paper we follow the relatively few studies that do consider the combination of different support schemes (Haegeland and Moen 2007a; Busom et al. 2011; Czarnitzki and Lopes-Bento 2014). This paper aims at assessing the effects of the federal tax incentives for R&D. However, given the importance of the subsidies provided by regional agencies, the total amount of subsidies received by companies is also included in the baseline specification as a control variable but also to evaluate the degree of complementarity or substitution between direct support and tax benefits (see section 3.3 on the Policy Mix). In the interpretation of the estimates of the direct support variable it should be kept in mind that, as the total amount of subsidies received by companies is considered, the variable on direct support does not account for the potential different effects of specific programmes and possible differences across regions.

The simplest estimation of a panel is so-called pooled Ordinary Least Squares (OLS) in which data on firms are pooled assuming that there is no heterogeneity across firms in the coefficients nor in the level of the dependent variable. This is a rather strong assumption, often rejected by the data, and most panel estimations account for firm-level heterogeneity by considering fixed or random effects. A fixed effects estimation considers firm-specific dummies that are constant over time.[16] In the context of public support for R&D, fixed effects can, as already pointed out by Lichtenberg (1984), represent unobserved firm characteristics that influence the firm's demand for public support, as long as these characteristics do not change much over the period considered. Henningsen et al. (2015) argue that firm fixed effects will capture firm characteristics such as R&D experience, experience with the application for public support and firm-specific technological opportunities. Using data on the evaluation of subsidy proposals in Norway, they find that unobserved project quality is largely absorbed by firm fixed effects. In a fixed effects specification the firm fixed effects are allowed to be correlated with the explanatory (control) variables,

[15] A traditional regression (Ordinary Least Squares, OLS) will only provide unbiased estimates if the assumptions with regard to the error term (sometimes labelled as disturbance or residual term) hold. With real observational data, the strong assumptions (e.g., homoscedasticity and no serial correlation) are often violated. Procedures that relax the assumptions need to be considered to account for the possible bias in the estimates.

[16] A fixed effects estimator is often called a within estimator as fixed effects estimates can also be obtained by OLS on variables after within transformation (subtracting, from each variable (dependent and independent), the average for each firm). The within estimation shows that this estimation considers changes over time within units (firms) rather than differences across (between) firms.

which can tackle time-invariant endogeneity of the explanatory (control) variables (Cameron and Trivedi 2010: p. 237).

A disadvantage of a fixed effects estimation is that time-invariant explanatory variables cannot be considered. Time-invariant dummies for firms with a small number of observations also tend to reduce the efficiency of estimates due to collinearity.

An alternative to fixed effects is a random effects specification, which assumes that there is no correlation between the firm-specific effects and the explanatory variables so that, in contrast with a fixed effects specification, time-invariant explanatory variables can be included. A Hausman test permits to decide whether random effects should be preferred to fixed effects.

Fixed and random effects specifications assume that firm heterogeneity is only reflected in level differences (differences in the intercept) and not in differences in the slope coefficients. Allowing for firm-specific coefficients is in most short panels not feasible as there are only a small number of observations for each firm.

The results of the estimation of the baseline panel specification (1) are reported in Table 7. The second column shows the results of a fixed effects (within) estimation that includes year dummies. Region and industry dummies, being fixed over time cannot be included in a fixed effects estimation. The third column shows the results of a fixed effects estimation in which industry x year dummies are also included to absorb possible industry-specific time-variant effects.

Table 7 Results of the baseline panel estimation (2003-2011)

	Fixed effects 1	Fixed effects 2	Random effects
Dependent variable (R&D expenditures net of public support)			
Explanatory variables (public support):			
Regional subsidy	0.07 (4.84)***	0.07 (4.97)***	0.15 (15.53)***
Research cooperation	0.09 (2.89)***	0.08 (2.44)**	0.11 (6.56)***
Young Innovative Company	-0.00 (-0.03)	0.01 (0.30)	0.20 (9.75)***
PhDs and civil engineers	0.06 (3.53)***	0.05 (2.45)***	0.17 (14.23)***
Master	0.11 (5.42)***	0.10 (4.29)***	0.12 (9.57)***
Tax credit R&D	-0.00 (-0.18)	0.00 (0.10)	0.03 (1.21)
Tax deduction 80% patent income	-0.00 (-0.10)	-0.03 (-0.95)	0.02 (0.72)
Control variables:			
Cash flow	0.11 (1.37)	0.12 (1.58)	0.24 (4.49)***
Number of employees	1.16 (3.85)***	1.20 (3.89)***	0.30 (3.57)***
Age	-0.07 (-0.68)	-0.10 (-0.98)	-0.01 (-3.10)**
Capital intensity	0.23 (1.73)*	0.21 (1.61)*	0.07 (1.28)
Region dummies	No	No	Yes
Industry (two-digit NACE)	No	No	Yes
Year dummies	Yes	Yes	Yes
Industry x year dummies	No	Yes	No
Number of observations	8,915	8,902	8,902

Note: The table shows the results of the estimation of the baseline panel specification (1) as shown in the text. All variables are considered in logs except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity. A robust Hausman test rejects random effects in favour of fixed effects at the 1% significance level.

Both estimations provide similar results indicating a positive impact of regional subsidies and three schemes of partial exemption from advance payment of the withholding tax on the wages of researchers (R&D cooperation; PhDs and civil engineers; Master degree). For the partial exemption scheme for Young Innovative Companies, the tax credit for R&D investment and the tax deduction of 80% of patent income, the effect of support on own R&D expenditures of firms is not found to be statistically significant. A random effects specification, reported in the last column in Table 7, provides larger coefficient estimates but also a statistically significant estimate for Young Innovative Companies and the tax credit for R&D investment, although the latter effect is only significant at 10%. A (robust) Hausman test indicates that a random effects specification is clearly rejected in favour of fixed effects.[17]

The number of employees of a company appears to be the most robust control variable, with an expected positive sign. The coefficients of cash flow and firm age also have the expected sign but are only statistically significant in the random effects specification. Capital intensity is positively correlated with R&D expenditures but the effect is only statistically significant at 10% in the fixed effects specification.

As pointed out in chapter 2 and shown in table A2.2 in Annex 2, due to a substantial number of missing values, the panel based on the matched data is highly unbalanced. Only for a rather small number of firms all variables are available over the entire period 2003-2011.

Although most estimation procedures can cope with unbalanced panels, estimates could be affected by the large number of firms with few observations (on panel attrition bias see, for example, Cameron and Trivedi 2010; Cheng and Trivedi 2014). To test the robustness of the results reported in Table 7, Table 8 shows the results of a fixed effects estimation, including year and industry x year dummies, for a (more) balanced panel. The second column shows the results for a fully balanced panel, considering only those firms that have no missing values for any year whereas the third column shows the results for a panel of firms with no missing values for at least five years. In the fully balanced panel, the positive effect of subsidies on R&D expenditures is statistically significant although only at the 10% level. The negative coefficient for Young Innovative Companies is also statistically significant at the 10%. None of the other coefficients of public support or of the control variables is significant.

The panel with firms with at least five years of full data contains substantially more observations and provides results that are more in line with the results reported in Table 7, although the coefficient estimates are smaller and the coefficient of the partial exemption for R&D cooperation is not statistically significant.

In the following subsections of 3.1, the results of some extensions and robustness tests of the baseline specification are reported.

---

[17] The Hausman test considers the null hypothesis that individual (firm) effects are random. If the random effects assumption holds, there should be no differences between a fixed effects or a random effects estimator. If so, a random effects is to be preferred as it is more efficient than fixed effects. If the null hypothesis is rejected, a fixed effects estimator is to be preferred as random effects and pooled OLS will not provide consistent estimates. However, the standard Hausman test assumes that there is no heteroskedasticity or serial correlation, an assumption that is violated in most panels. A robust Hausman test (RHAUSMAN in Stata, see Kaiser (2014)) can be used as an alternative that provides a robust test of the null hypothesis of a random effects specification. Both the standard and the robust Hausman test clearly reject the random effects null hypothesis (p-value=0.000).

Table 8 Results of balanced panel estimation (2003-2011)

	Fully balanced	At least five years of observations
Dependent variable (R&D expenditures net of public support)		
Explanatory variables (public support):		
Regional subsidy	0.07 (1.77)*	0.07 (3.80)***
Research cooperation	0.13 (1.22)	0.03 (0.90)
Young Innovative Company	-0.12 (-1.72)*	0.03 (0.41)
PhDs and civil engineers	-0.06 (-1.30)	0.04 (2.12)**
Master	-0.01 (-0.27)	0.05 (2.00)**
Tax credit R&D	0.05 (0.52)	-0.01 (-0.18)
Tax deduction 80% patent income	-0.06 (-1.46)	0.01 (0.27)
Control variables:		
Cash flow	0.12 (0.79)	0.07 (0.74)
Number of employees	0.98 (1.58)	0.79 (1.89)*
Age	-0.09 (-0.48)	0.02 (0.19)
Capital intensity	0.30 (0.96)	-0.03 (-0.15)
Region dummies	No	No
Industry (two-digit NACE)	No	No
Year dummies	Yes	Yes
Industry x year dummies	Yes	Yes
Number of observations	571	4,041

Note: The table shows the results of the estimation of fixed effects (within) estimation on a balanced panel. The second column reports the results of a fully balanced panel, which implies that there are no missing values for any of the variables over the entire period (nine years). The third column reports the results of an estimation for firms with no missing values for at least five out of nine years. All variables are considered in logs except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

3.1.1. Persistence of R&D activities and public support

Busom et al. (2015) point at recent empirical studies that acknowledge the very high persistence in R&D activities of companies. Possible explanations for this persistence are fixed costs of entry as well as exit from R&D activities which result in state dependence (Máñez et al. 2009; Arqué-Castells and Mohnen 2015); competition (Woerter 2014) or learning effects (Geroski et al. 1997; Triguero et al. 2014). Huergo and Moreno (2011), using data over the period 1990-2005, find that 89.1% of Spanish manufacturing firms that perform R&D activities in one year also perform R&D in the following year whereas 93.2% of firms that do not perform R&D in a given year also do not have R&D activities in the next year. The persistence of performers (non-performers) is lower (higher) for SMEs than for large firms, with respectively 83.2% persistence of R&D activities for SMEs, against 92.7% for large firms, and 95.1% persistence of no R&D activities for SMEs against 84.5% for large firms. Busom et al. (2015), also using data on Spanish manufacturing firms, report a persistence of 96% for firms that do not perform R&D and 90% persistence for R&D active firms over the period 2001-2008.

Of all firms in the dataset, for which data are available, 95.99% of firms with no R&D activities in a given year over the period 2003-2011 did not perform R&D in the next year and 92.25% of firms that did R&D in a given year also had R&D activities in the following year.

Table 9 shows the transition probabilities[18] of the status of R&D activities over the period 2003-2011 based on the Policy Mix R&D dataset. The results are very similar to those reported by Huergo and Moreno (2014) and Busom et al. (2015). Of all firms in the dataset, for which data are available[19], 95.99% of firms with no R&D activities in a given year over the period 2003-2011 did not perform R&D in the next year and 92.25% of firms that did R&D in a given year also had R&D activities in the following year.

Table 9 Persistence of R&D activities (2003-2011)

		Year t		
All firms		No R&D	R&D	Number of observations
Year t-1	No R&D	95.99	4.01	8,177
	R&D	7.75	92.25	6,319
	Number of observations	8,339	6,157	
SMEs (<= 250 employees)		No R&D	R&D	Number of observations
Year t-1	No R&D	96.67	3.33	6,811
	R&D	8.13	91.87	5,201
	Number of observations	7,007	5,005	
Large firms (> 250 employees)		No R&D	R&D	Number of observations
Year t-1	No R&D	94.48	5.52	851
	R&D	5.58	94.42	860
	Number of observations	852	859	

Note: The table shows the (transition) probabilities of firms changing from a status of R&D activities (No R&D vs. R&D) in a given year (t-1) to another status in the following year (t) for the period 2003-2011. Only real answers to the R&D survey are considered. Taking, for example, the first line of firms that did not have any R&D expenditures in year t-1, they have a probability of 95.99% not to have any R&D expenditures in the following year and a probability of 4.01% to have some R&D expenditures. Both rows and columns sum to 100.

Similar to Huergo and Moreno (2014), persistence of no R&D activities is higher, and persistence of R&D activities lower, for SMEs than for large firms although the degree of persistence exceeds 90% in all cases. Not too surprisingly, public support for R&D is as persistent as R&D activities. Busom et al. (2015) report high persistence of public support for R&D in Spain, distinguishing between subsidies and tax credits.

18 Transition probabilities show the probability that a firm with a given status in year *t-1* – for example, the firm does not perform any R&D – will have a given status – for example, that it will perform R&D – in year *t*.

19 Only real responses of firms on the R&D survey are reported, i.e. only this firms that respond whether they had any internal R&D activities in a given year or not.

Table 10 Persistence of public support for R&D (2003-2011)

		Year t				
	All firms	No support	Subsidy - no tax benefit	Tax benefit - no subsidy	Subsidy and tax benefit	Number of observations
Year t-1	No support	**95.54**	2.44	1.76	0.26	22,799
	Subsidy - no tax benefit	43.01	**47.98**	4.57	4.44	744
	Tax benefit - no subsidy	15.51	0.37	**72.29**	11.84	2,450
	Subsidy and tax benefit	3.91	2.56	28.42	**65.11**	665
	Number of observations	22,509	939	2,395	815	
	R&D active firms	No support	Subsidy - no tax benefit	Tax benefit - no subsidy	Subsidy and tax benefit	Number of observations
Year t-1	No support	**83.30**	10.09	5.31	1.30	922
	Subsidy - no tax benefit	40.78	**50.49**	4.85	3.88	103
	Tax benefit - no subsidy	3.91	0.37	**77.09**	18.62	537
	Subsidy and tax benefit	1.19	0.79	20.63	**77.38**	252
	Number of observations	834	149	520	311	834
	SMEs (<= 250 employees)	No support	Subsidy - no tax benefit	Tax benefit - no subsidy	Subsidy and tax benefit	Number of observations
Year t-1	No support	**95.62**	2.40	1.72	0.26	18,341
	Subsidy - no tax benefit	50.45	**40.43**	4.83	4.29	559
	Tax benefit - no subsidy	6.65	0.47	**80.21**	12.67	1,713
	Subsidy and tax benefit	3.56	2.57	29.64	**64.23**	506
	Number of observations	17,951	688	1,866	614	
	Large firms (> 250 employees)	No support	Subsidy - no tax benefit	Tax benefit - no subsidy	Subsidy and tax benefit	Number of observations
Year t-1	No support	**94.59**	1.71	3.70	0.00	702
	Subsidy - no tax benefit	40.91	**50.00**	0.00	9.09	22
	Tax benefit - no subsidy	4.27	0.36	**76.87**	18.51	281
	Subsidy and tax benefit	0.88	0.88	23.01	**75.22**	113
	Number of observations	686	25	268	139	

Note: The table shows the transition probabilities of firms changing from a status of public support in a given year (t-1) to another status in the following year (t). The period considered is 2009-2011 as in these years all benefits existed for at least one year. Taking, for example, the first line of firms that receive no public support (subsidy or tax benefit) for R&D in year *t-1*, they have a probability of 95.54% not to receive any support in the following year; a probability of 2.44% to receive a subsidy but no tax benefit; a probability of 1.76% to receive a tax benefit but no subsidy and a probability of 0.26% to receive a subsidy and at least one tax benefit. Both rows and columns sum to 100.

Table 10 shows transition probabilities of the status of support for the Policy Mix R&D data.

Four possibilities are considered for a given year:

– Firm does not receive any support for R&D (subsidy or tax benefit)
– Firm receives a subsidy but no tax benefit
– Firm receives at least one tax benefit but no subsidy
– Firm receives a subsidy as well at least one tax benefit.

For tax benefits, all four schemes of partial exemption from advance payment of the withholding tax are considered as well as the tax credit for R&D investment and the tax deduction of 80% of gross patent income. The transition probabilities are shown for all firms, for R&D active firms, for SMEs and for large firms. Only the period 2009-2011 is considered in the computation of the transition probabilities, as for these years all schemes of tax benefits existed for at least one year. Considering, for example, the first line, 95.54% of firms that did not receive a subsidy or a tax benefit in a given year did not receive any support in the following year; 2.44% received a subsidy but not tax benefit in the following year; 1.76% received a tax benefit but no subsidy and only 0.26% received a subsidy as well as at least one tax benefit in the following year.

The transition probabilities in Table 10 are again rather consistent with the probabilities reported in Busom et al. (2015), despite the differences in the tax benefits.[20] For example, the 95.54% of firms that persistently do not receive any support compares to the 95.9% for the panel of Spanish firms; the share of firms that persistently only receive a subsidy of 47.89% for Belgium compares to 59.7% for Spain; the share of firms that persistently only receive a tax benefit of 72.29% for Belgium compares to 60.4% for Spain and the share of firms that persistently receive direct support as well as at least one tax benefit of 65.11% for Belgium compares to 68.5% for Spain.

Table 11 shows the correlation over time of R&D expenditures and public support for R&D. The correlation over time of R&D expenditures is very high. Even with a three year lag, the correlation is still 0.93. Also for public support the correlation in t-3 is rather high, especially for subsidies and the partial exemption for researchers with a PhD or civil engineer degree.

The persistence of R&D activities and public support matters, both from a methodological perspective and a policy perspective. From a methodological point of view, estimates of the effectiveness of public support could be substantially biased if the high persistence is not accounted for. Busom et al. (2015) find that Spanish firms with finance constraints are less inclined to apply for tax credits. Persistence in tax credits could imply that mistakes in the design and implementation of the tax credit program are sustained, resulting in negative welfare effects. Akcigit and Kerr (2010) and Akcigit et al. (2014) argue that public support that is not selective, may result in overinvestment in applied research. The impact of public support on the orientation of R&D activities in Belgium is discussed in section 4.1.

---

[20] The only tax benefit in Spain is a tax credit whereas the tax benefits for Belgium consist in four schemes of partial exemption of advance payment of withholding tax, the tax credit for R&D investment and the tax deduction of 80% of gross patent income.

Table 11 Correlation over time of R&D expenditures and public support

	Year t-1	Year t-2	Year t-3
Year t			
R&D expenditures	0.98	0.95	0.93
Regional subsidy	0.74	0.71	0.66
Research cooperation	0.87	0.62	0.47
Young Innovative Company	0.85	0.67	0.48
PhDs and civil engineers	0.94	0.82	0.69
Master	0.89	0.70	0.51
Tax credit R&D	0.88	0.68	
Tax deduction 80% patent income	0.67	0.27	

Note: The table shows the correlation between the variables and the lagged variables in t-1 up to t-3. All correlations differ significantly from zero at the 5% significance level. The correlations for R&D expenditures and subsidies apply to the period 2003-2011; for the four schemes of partial exemption to the period 2008-2011 and for the tax credit and tax deduction of 80% of gross patent income to the period 2009-2011.

As pointed out by Zúñiga-Vicente et al. (2012), in a recent survey on public support for private R&D, most studies assess the contemporaneous impact, in effect, the effect of support in a given year on the R&D activities in the same year. They argue that as R&D projects take time to implement and can be subject to substantial adjustment costs, the impact of public support may be distributed over a longer period of time. Cerulli (2010) proposes a specification to which lags of public support are added.

Table 12 presents the results of the baseline specification to which a one year lag is added for all schemes of public support. The results for the contemporaneous effects of public support are in line with the results of the baseline specification in terms of statistical significance although the coefficients tend to be smaller.

With the one year lag, only the coefficient of regional subsidies is statistically significant. Given the high correlation over time of public support, adding lags introduces substantial collininearity. A test on the joint significance of the coefficients at the 1% level of public support indicates statistical significance for the coefficients of subsidies and the partial exemption for researchers with a master degree and statistical significance at the 10% level for the coefficients of the partial exemption for researchers with a PhD or civil engineer degree.

The results of a specification in which two year lags are included are reported in Table 13. Only the contemporaneous effect of subsidies and the partial exemption for researchers with a PhD or civil engineer degree are statistically significant positive, though only at the 10% significance level. For the one year lag effects, only the positive coefficient of Young Innovative Companies is significant (just above 5% significance level) and for the two year lag only the positive coefficient of the partial exemption for researchers with a master degree is statistically significant.

Table 12 Results of alternative panel specifications with one lag for public support (2003-2011)

	Contemporaneous	One year lag
Dependent variable: (R&D expenditures net of public support)		
Explanatory variables (public support):		
Regional subsidy	0.05 (2.90)***	0.05 (2.99)***
Research cooperation	0.06 (1.77)*	0.01 (0.47)
Young Innovative Company	0.06 (0.81)	0.01 (0.24)
PhDs and civil engineers	0.04 (2.18)**	-0.00 (-0.14)
Master	0.06 (2.53)***	0.03 (1.32)
Tax credit R&D	0.02 (0.67)	-0.02 (-0.60)
Tax deduction 80% patent income	0.00 (0.17)	-0.01 (-0.51)
Control variables:		
Cash flow	0.05 (0.55)	
Number of employees	0.79 (1.80)*	
Age	-0.03 (-0.26)	
Capital intensity	0.06 (0.39)	
Number of observations: 5,924		

Note: The table shows the results of the estimation of the baseline specification in which a lag is introduced for public support. All estimations use fixed effect (within) with industry x year as well as year dummies.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

Table 13 Results of alternative panel specifications with two lags for public support (2003-2011)

	Contemporaneous	One year lag	Two year lag
Dependent variable: (R&D expenditures net of public support)			
Explanatory variables (public support):			
Regional subsidy	0.03 (1.93)*	0.01 (0.41)	0.01 (0.69)
Research cooperation	0.06 (1.45)	-0.01 (-0.65)	0.02 (0.75)
Young Innovative Company	0.00 (0.06)	0.08 (1.94)*	-0.03 (-0.77)
PhDs and civil engineers	0.03 (1.80)*	-0.00 (-0.27)	0.01 (0.59)
Master	0.02 (0.86)	0.01 (0.59)	0.04 (2.12)**
Tax credit R&D	0.03 (1.21)	-0.01 (-0.42)	0.02 (0.91)
Tax deduction 80% patent income	-0.02 (-1.17)	0.02 (0.82)	0.00 (0.30)
Control variables:			
Cash flow		0.20 (2.08)**	
Number of employees		0.29 (0.49)	
Age		-0.24 (-1.63)	
Capital intensity		-0.08 (-0.45)	
Number of observations: 3,447			

Note: The table shows the results of the estimation of alternative panel specifications, using the baseline specification in which tow lags are introduced for public support. All estimations use fixed effect (within) with industry x year as well as year dummies.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

A test of the joint significance of the three coefficients for each scheme of support only results in the rejection (marginally at 10%) of the null hypothesis of no significance for the partial exemption for researchers with a master degree. By including two lags, the number of observations drops to 3,447 and the specification is even more likely to suffer from substantial collinearity.

An alternative specification to assess the potential lagged impact is to replace contemporaneous support by lags. Table 14 reports the results of a panel estimation in which, respectively a one and a two year lag is used for each support scheme. The results of the one year lag specification are again in line with the results of the baseline specification. In the two year lag specification, only the positive effect of the partial exemption for researchers with a master degree is statistically significant. The very high correlation over time, as shown in Table 11, implies that persistence in firm-level R&D expenditures needs to be accounted for in the evaluation of public support for R&D.

Table 14 Results of alternative panel specifications with two lags (2003-2011)

	One year lag	Two year lag	Lagged dependent	Lagged dependent + Lagged support
Dependent variable (R&D expenditures net of public support)				
Explanatory variables:				
Net R&D expenditures (t-1)			0.48 (26.69)***	0.48 (26.75)***
Regional subsidy	0.05 (3.16)***	0.01 (0.74)	0.01 (0.77)	0.04 (3.23)***
Research cooperation	0.05 (1.71)*	0.01 (0.66)	0.02 (1.28)	-0.00 (-0.14)
Young Innovative Company	0.03 (0.76)	0.02 (0.43)	0.03 (0.74)	0.02 (0.76)
PhDs and civil engineers	0.03 (1.73)*	0.02 (0.95)	0.03 (2.20)**	0.01 (1.14)
Master	0.07 (2.72)***	0.05 (2.39)**	0.03 (1.86)*	0.03 (2.08)**
Tax credit R&D	-0.01 (-0.49)	0.01 (0.61)	-0.00 (-0.18)	-0.02 (-1.07)
Tax deduction 80% patent income	-0.01 (0.23)	0.03 (0.98)	0.00 (0.02)	0.01 (0.43)
Control variables:				
Cash flow	0.07 (0.74)	0.20 (2.19)**	0.06 (0.71)	0.05 (0.71)
Number of employees	0.87 (1.99)**	0.30 (0.55)	0.50 (1.58)	0.52 (1.63)*
Age	0.00 (0.01)	-0.21 (-1.49)	0.43 (2.75)***	0.44 (2.87)***
Capital intensity	0.07 (0.45)	-0.02 (-0.12)	0.03 (0.25)	0.03 (0.31)
Region dummies	No	No	No	No
Industry (two-digit NACE)	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Industry x year dummies	Yes	Yes	Yes	Yes
Number of observations	5,924	3,554	5,770	5,770

Note: The table shows the results of the estimation of alternative panel specifications, using lagged variables. All variables are considered in logs except firm age. All estimations use fixed effect (within) with industry x year as well as year dummies.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

The most straightforward way to account for persistence in the dependent variable, is to include its lag as an explanatory variable. However, including a lagged dependent variable poses a number of estimation problems. If a lagged dependent variable is included in a model that contains serial correlation, estimates from OLS and fixed effects will be inconsistent (e.g. Greene 2000; Cameron and Trivedi 2010). Achen (2001) argues that in the presence of serial correlation, a lagged dependent variable tends to dominate and suppress the effects of other explanatory variables. According to Keele and Kelly (2006) the use of a lagged dependent variable can be appropriate in some circumstances but, for example, not when data are not stationary (see section 3.9). Wilkins (2014), on the other hand, states that excluding a lag of the dependent variable may lead to substantially biased estimates and that the inclusion of the lagged dependent variable should be part of a robustness test of estimation. As a first indication of the potential impact of persistence in R&D expenditures on the estimation, Table 14 reports the results of a panel estimation in which the lagged level of R&D expenditures is included.

The fourth column shows the results of including the lagged dependent in the baseline specification and the fifth column the results of a specification in which the lagged dependent is included in a specification with one year lags for public support. The large and highly significant coefficient confirms the strong persistence in firm-level R&D. In the contemporaneous specification with a lagged dependent variable, the estimate of the effect of the partial exemption for researchers with a PhD and civil engineer degree and researchers with master degree are statistically significant, though the latter only at 10%. In the one-year lag specification the positive coefficients of subsidies and the partial exemption for researchers with a master degree are significant.

Cerulli (2010) points out that if public support is not exogenous or predetermined, a specification with lagged support variables may be problematic even if contemporaneous support is not included in the regression. If the error term is serially correlated[21], the lags of support will also be endogenous and instrumental variable estimation is necessary.

The results of the specifications using lags should be interpreted with caution given the issues previously mentioned. The inclusion of lags of the dependent and independent variables is further discussed in the section on dynamic panel estimation (section 3.9).

3.1.2. Policy Mix

Table 5 in Chapter 1 shows the extent to which firms combine different schemes of public support for R&D. The most common combination is the partial exemption from withholding tax on the wages of researchers with a PhD or civil engineer degree with the partial exemption for researchers with a master degree. A large share of firms combine more than two support schemes, especially firms that use the tax credit for R&D investment and the tax deduction of 80% of gross patent income.

Busom et al. (2015) point out that most studies on public support for R&D assess the impact of individual support schemes although in most countries firms can receive direct support as well as tax benefits. They argue that ignoring the possible combination of different schemes may result in biased estimates of public support. Hægeland and Møen (2007a) were among the first to assess the interaction between subsidies and a R&D tax credit in Norway. Their results suggest that direct and indirect support are complements. Czarnitzki and Lopes-Bento (2014) also find indications of complementarity between R&D subsidies granted by the German government and research support by the European Commission. The baseline panel specification estimates the impact of each support scheme conditionally on the other support schemes, avoiding the possible bias due to omitting available information on other forms of

[21] A Wooldridge panel test for autocorrelation clearly rejects the null hypothesis that there is no first-order autocorrelation. In a specification that accounts for a first-order autoregressive error term the coefficients of subsidies and the partial exemption for R&D cooperation are not statistically significant whereas the coefficient of the partial exemption for researchers with a PhD or civil engineer degree and researchers with a Master degree are only significant at 10%. On the other hand, a fixed effects estimation with standard errors that are robust to heteroskedasticity, autocorrelation up to lag 4 as well as cross-sectional dependence, following Driscoll and Kraay (1998), confirms the statistical significance of regional subsidies and three of the four schemes of partial exemption of the fixed effects specifications reported in table 7. The Wooldridge test is performed using XTSERIAL, the regression with a first-order autoregressive error term using XTREGAR and the fixed effects regression with robust standard errors using XTSCC, all in Stata. As the XTSCC estimator is based on asymptotic theory, Hoechle (2007) points out that caution is warranted in applying this estimator to panels that contain a large cross-section but only a short time dimension, as is the case in our panel.

public support. However, the specification does not account for possible complementarities or substitution that may result from the combination of different support schemes. Table 15 shows the results of an estimation of the baseline specification to which variables have been added that reflect the support received by firms that combine two forms of support. Given seven different schemes, 21 meaningful combinations exist. Only the combinations for which the coefficient is statistically significant are shown. The coefficients of the combination variables should be interpreted as additional effects of combining two schemes relative to the individual schemes. A positive coefficient implies that combining the two schemes provides a complementary impact and a negative coefficient would indicate that the combination results in some crowding out of R&D expenditures.

The statistical significance of the coefficients of the individual support schemes corroborate the estimates of the baseline specification but the coefficients are generally higher. This seems to be explained by indications of substitution between the different schemes. Five out of the seven combinations of two schemes for which the effect is statistically significant, are negative. Only the combination of the partial exemption for Young Innovative Company and the tax deduction of 80% gross patent income and the combination of the partial exemption for researchers with a master degree and the tax credit have a positive coefficient. Although the additional effect of seven combinations is statistically significant, the coefficients are very small. This is however not true for the coefficient of firms that combine more than two support schemes. The negative coefficient is statistically significant but also considerable. These results suggest that the positive impact of subsidies and some of the tax benefits is reduced by the combination of different schemes, especially when more than two schemes are combined.[22]

[22] A similar conclusion was found in the first evaluation of the Belgian tax incentives for R&D (see Dumont 2013).

Table 15 Results of the estimation of the Policy Mix of public support for R&D (2003-2011)

Dependent variable (R&D expenditures net of public support)	
Explanatory variables (individual support scheme):	
Regional subsidy	0.08 (5.56)***
Research cooperation	0.13 (3.82)***
Young Innovative Company	0.03 (0.68)
PhDs and civil engineers	0.09 (5.11)***
Master	0.14 (6.47)***
Tax credit R&D	0.01 (0.36)
Tax deduction 80% patent income	0.02 (0.79)
Explanatory variables (combination support):	
Regional subsidy + Master	-0.00 (-2.15)**
Research cooperation + PhD and civil engineers	-0.00 (-4.35)***
Research cooperation + Master	-0.00 (-1.98)**
Young Innovative Company + Master	-0.00 (-6.37)***
Young Innovative Company + Tax deduction 80% patent income	0.00 (4.56)***
Master + Tax credit R&D	0.00 (11.90)***
Master + Tax deduction 80% patent income	-0.00 (-2.33)**
More than two schemes combines	-0.07 (-4.13)***
Number of observations	8,915

Note: The table shows the results of an estimation of the baseline specification to which terms are added which reflect the amount received by firms that combine two support schemes. All variables are considered in logs except firm age. The four control variables are included in the estimation but not reported. All estimations use fixed effect (within) with year dummies.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

3.1.3. Alternative dependent variables

In previous studies on public support for R&D, some alternatives to the (log) level of R&D expenditures are considered as the dependent variable:

- R&D intensity, mostly R&D expenditures over sales (for example, Meeusen and Janssens 2001; Aerts and Czarnizki 2004; Gonzales et al. 2005)
- R&D expenditures per employee (for example, Busom 2000; Lööf and Heshmati 2005; Görg and Strobl 2007; Hussinger 2008)
- The number of researchers or total R&D personnel (for example, Suetens 2002; Wolff and Reinthaler 2008)
- Researchers as a share of total personnel (for example, Globerman 1973; Busom 2000)

As sales are not included in the data, value added is used as the denominator for R&D intensity, with the possible disadvantage that value added can be negative in which case the observation is disregarded in the estimation. Sales (turnover) is a firm-level variable that is also available from Bureau Van Dijk, but not included in the matched data. A limitation of using sales for Belgium is that companies that are allowed to deposit an annual account in abbreviated rather than in full form do not have to report sales.

Eligibility to deposit an annual account in abbreviated form is based on size.[23] Using sales as denominator for R&D intensity would therefore bias results toward large companies. An additional problem is that some young R&D intensive firms do not have any sales and are therefore disregarded even if they would be required to report turnover. A more fundamental problem with R&D intensity is that it is a ratio of R&D input relative to some indicator of output. Changes in intensity result from changes in the nominator as well as in the denominator and using R&D intensity as the dependent variable may confound different effects. Coad and Rao (2010) provide evidence for US manufacturing firms, covering the period 1974-2004, that firms appear to aim at constant R&D intensity (defined as R&D to employment or to sales), increasing their R&D expenditures following growth in sales and employment. This finding is in line with previous indications that firms tend to increase their R&D expenditures proportionally with firm size (for example, Cohen and Klepper 1996; Symeonidis 1996; Klette and Kortum 2004). In some recent studies, a negative link is found between R&D intensity and firm size (for example, Akcigit 2008; Park 2010; Stancík and Biagi 2012; Busom et al. 2015). According to Park (2010), the link between R&D intensity and firm size may have changed over time with a new type of firm appearing in the 1980s, with high R&D expenditures but hardly any revenue. In the policy mix R&D data for Belgium the correlation between the ratio of R&D to value added and firms size (value added or number of employees) is slightly negative but not statistically significant. The link between R&D employment intensity (Researchers/employees) and firm size is also negative and statistically significant at the 5% level.

Lichtenberg (1984) pointed out that R&D expenditures are either considered in current prices (not deflated) or in constant prices, with expenditures deflated by some price index which is, given the well-known problems in constructing R&D deflators[24], likely to be a poor proxy of the marginal cost of R&D input. Using the same price index to deflate the dependent and the explanatory variables, will induce a spurious positive correlation whereas if the variables are not deflated the bias will depend on the elasticity of the demand for own R&D activities and public support.[25] Lichtenberg (1984) argues that given the difficulties in finding reliable deflators for R&D expenditures, data on R&D employment may provide a better indication of real R&D input than inaccurately deflated R&D expenditure. Two measures of R&D employment are considered, the total number of employees involved in R&D activities and the ratio of researchers to the total number of employees. The results of the four alternative dependent variables of R&D input are reported in Table 16.

Considering R&D expenditures relative to value added as the dependent variable, only the coefficient of regional subsidies is statistically significant. The sign is, in contrast with most other estimates, negative. The results of the other alternative measures of R&D input are more in line with the estimates of the baseline specification although the positive coefficient for the partial exemption for researchers with

---

23 A firm is regarded as small if it has not exceeded more than one of the following ceilings in the last two financial years for which the accounts are closed:
- annual average workforce: 50
- turnover (excluding VAT): 7,300,000 EUR
- balance sheet total: 3,650,000 EUR,
unless the annual average workforce exceeds 100 units. (Source: National Bank of Belgium website).

24 See for example, Haskel et al. (2011) and Robbins et al. (2012).

25 In this paper all variables expressed in euro have been delated using industry price deflators from FPS Economy. Estimations using all variables in current prices (not deflated) provide similar results and do not change the conclusions as to the sign or statistical significance of the estimates.

a master degree is not statistically significant in the two specifications with an employment-based indicator, suggesting that this measure has a stronger impact on R&D expenditures than on overall R&D employment. The impact of public support on R&D employment and the composition of R&D personnel is further analyzed and discussed in section 4.2.

The lack of evidence of the impact of public support on R&D intensity may seem worrisome as the rationale for the federal tax benefits relies on their contribution to the 3% R&D intensity target.

Table 16 Results of panel estimation with alternative dependent variables (2003-2011)

Dependent variable:	$\frac{R\&D\ expenditures}{value\ added}$	$\frac{R\&D\ expenditures}{Number\ of\ employees}$	*Number of researchers (FTE)*	$\frac{R\&D\ personnel}{Number\ of\ employees}$
Explanatory variables:				
Regional subsidy	-0.02 (-4.57)***	0.13 (5.82)***	0.08 (5.53)***	0.09 (5.87)***
Research cooperation	-0.00 (-0.05)	0.17 (3.09)***	0.11 (3.15)***	0.11 (3.16)***
Young Innovative Company	-0.03 (-1.14)	-0.08 (-0.59)	0.04 (0.65)	0.05 (0.76)
PhDs and civil engineers	0.01 (1.47)	0.09 (2.43)**	0.06 (2.43)**	0.06 (2.36)**
Master	-0.00 (-0.25)	0.08 (2.01)**	0.04 (1.50)	0.05 (1.54)
Tax credit R&D	0.01 (1.54)	0.05 (1.00)	0.03 (0.81)	0.03 (0.73)
Tax deduction 80% patent income	-0.00 (-0.51)	0.02 (0.49)	-0.01 (-0.22)	-0.01 (-0.32)
Control variables:				
Cash flow	-0.00 (-0.13)	0.17 (1.36)	0.07 (0.93)	0.07 (0.88)
Number of employees	0.34 (3.95)***	1.19 (2.32)**	1.68 (5.13)***	0.47 (1.39)
Age	1.15 (44.47)***	-0.31 (-1.71)*	-0.18 (-1.56)	-0.19 (-1.58)
Capital intensity	0.05 (1.47)	0.09 (0.38)	0.02 (0.15)	0.04 (0.26)
Region dummies	No	No	No	No
Industry (two-digit NACE)	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Industry x year dummies	Yes	Yes	Yes	Yes
Number of observations	7,423	9,396	8,730	8,730

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) using alternative dependent variables. All variables are considered in logs, except firm age and the dependent variables in the second and last column. As these are bounded between 0 and 1 a logistic transformation has been applied.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity

However, R&D intensity in employment terms, as reported in the last column of Table 16, does provide indications that some of the support schemes have a positive effect on R&D intensity, suggesting possible problems in deflating when using an intensity measure based on two monetary variables. Moreover, even if R&D intensity of firms is rather constant, as the evidence mentioned before indicates, aggregate R&D intensity can still increase if the share of R&D active firms and especially the share of R&D intensive firms in the economy increases (see section 3.10 on the effects on starting R&D).

3.1.4. Industry and firm heterogeneity

The baseline panel specification pools data on all firms across different industries. Time-invariant firm-level heterogeneity is captured by firm fixed effects and time-varying industry-level effects are reflected by the industry x year dummies. This specification relies on the rather strong assumption that there are no differences across firms, according to some firm characteristics, or across industries, according to some industry characteristics, in the impact of public support on firm R&D activities.

A recent survey by Castellacci and Lie (2015) shows that the effects of public support for R&D differ across industries. The survey – based on studies that mainly evaluate the effects of tax credits – concludes that additionality is higher for SMEs, in service sectors and somewhat surprisingly also in low-tech industries although this latter conclusion especially applies to countries with incremental rather than volume-based support schemes. Following the Pavitt classification, discussed further on in this section, Bodas Freitas et al. (2015) find stronger effects of tax credits in science-based and specialized supplier industries than in supplier-dominated industries. They argue that the evidence on cross-industry heterogeneity raises the question whether public support for R&D should not be differentiated by industry, for example by providing more support in industries with high R&D intensity and technological opportunities and consider other, more appropriate, instruments of innovation policy in industries where R&D is generally not the main innovation strategy. The fact that the conclusions with regard to R&D intensity of Castellacci and Lie (2015) diverge somewhat from those of Bodas Freitas et al. (2015) should however warn for hasty conclusions.

In this section, the baseline specification is estimated for sub-samples, grouping firms by a number of firm-level variables as well as by some industry characteristics.

Table A 3.2 in Annex 3 shows the results of an estimation of the baseline specification for individual industries, using the top 10 industries in terms of the number of firms that receive public support for R&D. Table A 3.1 lists the two-digit industries that appear in the dataset, with the number of observations for responses to the R&D survey. Somewhat surprisingly, the three industries with the highest number of observations are not manufacturing industries but *Computer programming, consultancy and related activities* (NACE 62); *Architectural and engineering activities; technical testing and analysis* (NACE 71) and *Wholesale trade, except of motor vehicles and motorcycles* (NACE 46). Especially the high number of firms with R&D activities and public support for R&D that are classified in the latter industry may indicate some problems of industry classification.[26] The results in table A3.2 show some heterogeneity in the effects of public support. The effect of subsidies is found to have a statistically significant positive effect in three industries and the impact of the partial exemption for researchers with a master degree in six industries. The coefficient of the tax credit for R&D is significantly positive in two industries but also significantly negative in *Manufacture of chemicals and chemical products* (NACE 20), which could explain the lack of a clear indications on the effect of the tax credit in the whole panel. In one industry in the top 10, *Manufacture of computer, electronic and optical products* (NACE 26), the coefficient of the tax deduction of 80% of gross patent income is actually also positive and statistically significant, though only at the 10% significance level.

---

[26] The industry classification of firms used in the analysis is from the Belfirst data (Bureau Van Dijk).

Although they provide indications of industry heterogeneity in the effects of public support, the estimations of individual industries should be considered with some caution. In some industries the number of observations is fairly low and some of the estimates are clearly not realistic, for example the statistically significant (at 10%) 8.10 elasticity of the partial exemption for Young Innovative Companies in *Manufacture of machinery and equipment n.e.c* (NACE 28) or the significant -1.01 elasticity of the partial exemption for Young Innovative Companies in *Activities of head offices; management consultancy activities* (NACE 70). Baltagi and Griffin (1997) provide evidence of the advantage of pooled panel estimators over heterogeneous models which often produce less plausible estimates and overall forecast performance. Moreover, finding different results across individual industries does not provide any insight as to the possible explanation for the different impact of public support. It is therefore more interesting to consider specific groups of firms or industries to assess whether the differences can be explained by firm or industry characteristics.

A number of studies assess possible different effects of public support for R&D by firm size, with mixed effects. Whereas some studies conclude that support appears to be more effective for small and medium-large firms (Lokshin and Mohnen 2007) others find stronger effects for large firms (Corchuelo and Martínez-Ros 2009; Cerulli and Poti 2010). Streicher, Schibany and Gretzmacher (2004) found that input additionality was higher for small and large companies than for medium-large companies in Austria. Alternative characteristics that are considered to group firms in this paper are R&D intensity and the updated version of the Pavitt category provided by Bogliacino and Pianta (2015).

In Table 17 the results are shown for separate estimations of the baseline specification for four groups, with firms classified by the number of employees. The results of the group of firms with at most 50 employees (by far the largest group) are most in line with the results of the full panel, with a statistically significant positive coefficient for regional subsidies and the schemes of partial exemption for researchers, except for Young Innovative Companies.

The effect of regional subsidies is also significantly positive for firms with a number of employees between 50 and 100 and a number of employees above 250. For the latter group the coefficient of the partial exemption for researchers with a PhD or civil engineer degree and researchers with a master degree is also significantly positive but only at the 10% significance level. In the group of the largest firms, the coefficient of the tax credit is negative and statistically significant though also only at 10%. The effect of the partial exemption for researchers involved in R&D cooperation is positive and significant, again only at 10%.

The lower statistical significance can to some extent be explained by the fact that the group of firms with at most 50 employees is far larger than the three other groups, in effect counts more observations than the three other groups combined. A strange result in Table 17 is that in the two groups with the largest firms (fourth and fifth column) the coefficient for Young Innovative Companies is actually estimated. This implies that some of the Young Innovative Companies have more than 100 and even more than 250 employees. Although this is not entirely impossible, the fact that a small number of these companies – according to the firm-level information – are older than 10 years at the time that they receive support under the Young Innovative Companies scheme (see Table 18) – which is, by definition, excluded – seems problematic although it may be due to inconsistencies in the linking of the different data sources.

Table 17 Results of panel estimation by firm size (2003-2011)

	employees<=50	50< employees<=100	100< employees<=250	250< employees
Dependent variable (R&D expenditures net of public support)				
Explanatory variables:				
Regional subsidy	0.05 (2.90)***	0.09 (2.08)**	0.04 (1.49)	0.07 (2.21)**
Research cooperation	0.16 (2.75)***	0.03 (0.56)	0.13 (1.76)*	0.02 (0.49)
Young Innovative Company	-0.03 (-0.76)	-	1.59 (0.26)	1.28 (13.35)
PhDs and civil engineers	0.08 (2.60)***	0.05 (1.23)	-0.04 (-0.95)	0.08 (1.87)*
Master	0.11 (3.52)***	0.02 (0.34)	0.02 (0.38)	0.07 (1.72)*
Tax credit R&D	0.04 (1.01)	-	-0.06 (-1.35)	-0.06 (-1.66)*
Tax deduction 80% patent income	-0.00 (-0.09)	0.01 (0.20)	0.02 (0.26)	-0.05 (-1.12)
Control variables:				
Cash flow	0.06 (0.50)	0.31 (1.31)	-0.03 (-0.23)	0.18 (0.97)
Number of employees	1.42 (4.40)***	1.90 (2.57)***	1.11 (1.40)	0.35 (0.38)
Age	-0.31 (-2.14)**	0.27 (0.87)	0.05 (0.24)	0.14 (0.47)
Capital intensity	0.42 (2.76)***	-0.19 (-0.46)	-0.40 (-0.81)	0.12 (0.22)
Region dummies	No	No	No	No
Industry (two-digit NACE)	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Industry x year dummies	No	No	No	No
Number of observations	5,133	1,262	1,167	1,352

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) by size category. All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

In Table 18 the results are shown of an estimation for groups classified by firm age. The age groups are not equally sized as the majority of firms in the panel are older than 10 years. The results for the group of firms older than ten years are therefore most in line with the results for the whole panel. The fact that the coefficient for the partial exemption for Young Innovative Companies is actually estimated for this group reveals that some of the Young Innovative Companies are older than 10 years, in the year they receive the support under this scheme, which is not possible. The finding does not necessarily imply irregularities as it could be due to data problems, with firm age defined by the date of creation in the Belfirst data.

Table 18 Results of panel estimation by firm age (2003-2011)

	age<=5	5< age<=10	10<age
Dependent variable (R&D expenditures net of public support)			
Explanatory variables:			
Regional subsidy	-0.07 (-2.45)**	-0.03 (-1.21)	0.09 (4.99)***
Research cooperation	0.06 (0.80)	0.01 (0.26)	0.08 (2.32)**
Young Innovative Company	-0.07 (-0.50)	0.05 (0.84)	0.66 (1.42)
PhDs and civil engineers	0.02 (0.33)	0.11 (2.65)***	0.05 (2.87)***
Master	0.14 (1.50)	0.02 (0.24)	0.08 (3.69)***
Tax credit R&D	-0.02 (-0.42)	0.07 (1.19)	0.02 (0.51)
Tax deduction 80% patent income	-0.14 (-2.19)**	0.05 (0.89)	-0.01 (-0.31)
Control variables:			
Cash flow	0.20 (1.11)	0.40 (1.95)**	0.13 (1.48)
Number of employees	0.37 (0.90)	-0.11 (-0.12)	1.15 (3.41)***
Age	0.13 (1.31)	-0.31 (-0.94)	-0.06 (-0.50)
Capital intensity	0.41 (1.31)	0.30 (0.85)	0.19 (1.16)
Region dummies	No	No	No
Industry (two-digit NACE)	No	No	No
Year dummies	Yes	Yes	Yes
Industry x year dummies	No	No	No
Number of observations	777	1,120	7,017

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) by firm age category. All variables are considered in logs, except firm age.

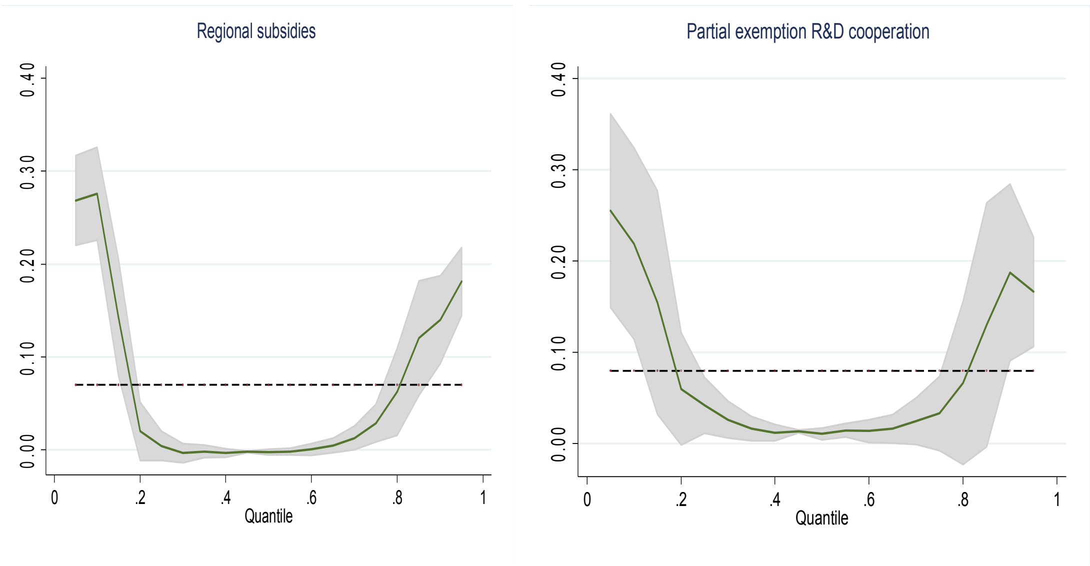
*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

The most surprising result is the statistically significant negative coefficient for regional subsidies as well as for the tax deduction of 80% of gross patent income in the group of firms younger than 5 years. As with the estimations by size category, the least significant results are found for the intermediate age group, for which only the effect of the partial exemption for researchers with a PhD or civil engineer degree is statistically significant positive.

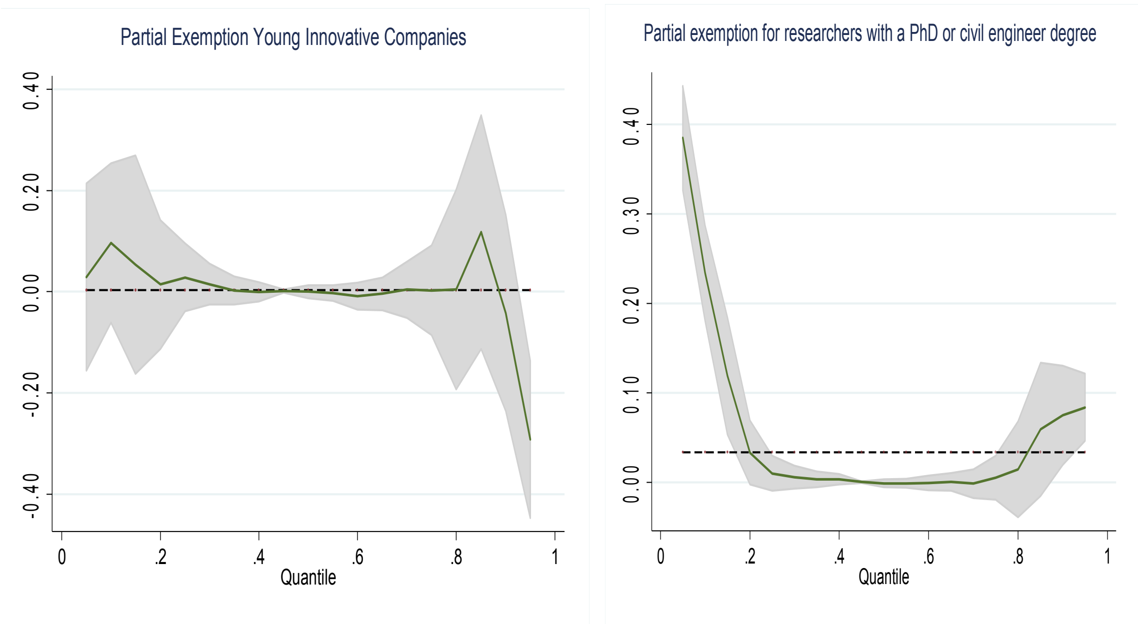
The result for firms classified by firm size and firm age indicate that the impact of subsidies and tax incentives is not necessarily constant across some firm characteristics. To assess potential non-linear effects of public support, graphs 6 up to 9 represent the coefficients of the public support variables from a quantile regression.[27] The coefficients reflect the quantiles of the conditional distribution as linear functions of the explanatory (right-hand-side) variables. The horizontal line denotes the OLS estimate, which provides an average estimate across the entire distribution. The grey-shaded areas are the 5% confidence intervals of the coefficient estimates.

[27] The graph result from running the Stata procedure GRQREG.

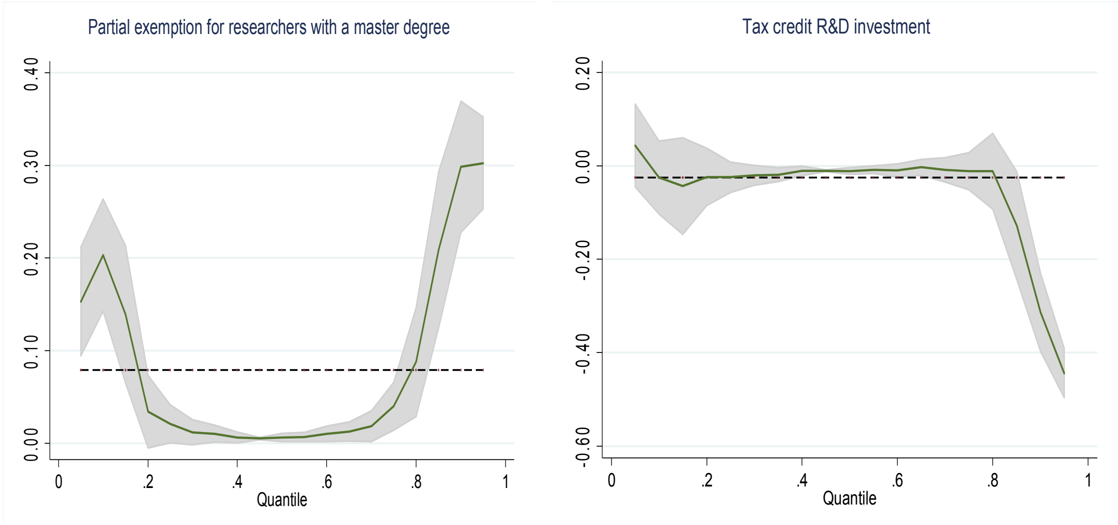
Graph 6 Impact by quantile: regional subsidies and partial exemption R&D cooperation



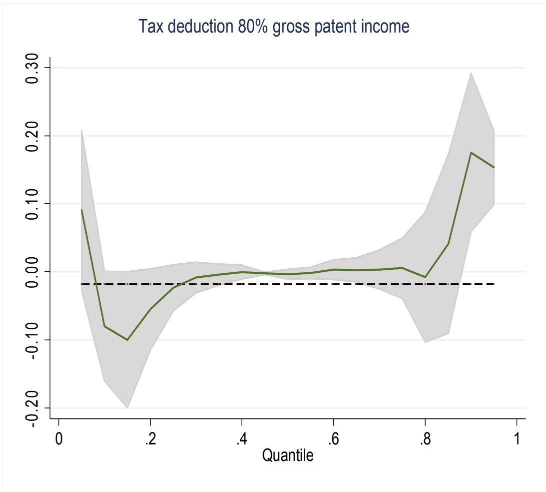
Graph 7 Impact by quantile: partial exemption YIC and researchers with a PhD or civil engineer degree



Graph 8 Impact by quantile: partial exemption researchers with a master degree and tax credit R&D



Graph 9 Impact by quantile: tax deduction of 80% of gross patent income



The impact of public support is not linear or monotonic for any of the schemes. For five schemes, the effect appears to be U-shaped, with a high elasticity at the low and high end of the distribution of support and a lower elasticity in the intermediate range. The confidence interval of estimates is clearly larger for those support schemes with relatively few observations (partial exemption for R&D cooperation and Young Innovative Companies and the tax deduction of 80% of patent income). For subsidies and the partial exemption for R&D cooperation the elasticity of R&D expenditures with respect to the amount of support is relatively equal at both ends of the distribution. The partial exemption for researchers with a PhD or civil engineer degree appears to result in additionality at the lower end of support whereas the partial exemption for researchers with a master degree is somewhat higher for higher amounts of support. For the three tax incentives for which little indications of any impact on private R&D expenditures are found in the estimation of the baseline specification, the coefficient estimate is close to zero over the entire distribution. For the partial exemption for Young Innovative Companies and the tax credit for R&D investment, the coefficient drops substantially at the higher end of support whereas for the tax deduction of 80% of patent income there is a positive and statistically significant impact at the upper end of support.

Instead of using firm characteristics to group firms, estimations can be done for groups of industries, categorized by some industry characteristic. A well-known classification of industries that is often used in R&D and innovation studies, is the categorization proposed by Pavitt (1984), based on the origin and main sources of technological knowledge and market structure of industries. Boggliacino and Pianta (2015) recently proposed an update of the Pavitt categories by NACE industries (Revision 2 - 2008), including a classification for new industries and services. The four Pavitt categories are:

1. **Science Based**: sectors in which innovation is based on advances in science and R&D.
2. **Specialized Suppliers**: sectors producing machinery and equipment that is used in new processes for other industries.
3. **Scale Intensive**: sectors in which scale economies are relevant and a certain rigidity of production processes exists, technological change is usually incremental.
4. **Supplier-dominated**: traditional sectors in which small firms are prevalent and technological change is introduced through the inputs and machinery provided by suppliers from other industries.

Table 19 shows the results of a fixed effects specification with industries grouped by the updated Pavitt classification provided by Boggliacino and Pianta (2015). The estimates of regional subsidies and the partial exemption for researchers with a master degree, are positive and statistically significant, though sometimes only at the 10% significance level, for all four Pavitt categories. The coefficient of the partial exemption for researchers with a PhD or civil engineer degree is positive and significant for the group of *Specialized Suppliers* and *Supplier-dominated* industries. In *Specialized Suppliers* industries the effect of the partial exemption for Young Innovative Companies is negative and statistically significant (10% level) and positive and statistically significant in *Scale Intensive* industries, though the latter estimate is unrealistically high. The results for Belgium are not in line with the results reported by Bodas Freitas et al. (2015), who find, in an analysis for France, Italy and Norway, that firms tend to respond more to tax incentives for R&D in science-based and specialized supplier industries than in supplier-dominated industries.

In Table 20 industries are grouped by the orientation of R&D activities. Industries are ranked by the average share that firms spend on respectively basic research, applied research and experimental development. Consequently, industries are split into three equally sized groups according to the predominance of the orientation of their R&D activities. In industries with a relatively high share of R&D spend on basic research, the effect of subsidies and the partial exemption for researchers with a PhD or civil engineer and for researchers with a master degree is positive and statistically significant although the coefficient of subsidies is only significant at the 10% level.

Table 19 Results of panel estimation by Pavitt category (2003-2011)

	Science Based	Specialized Suppliers	Scale Intensive	Supplier Dominated
Dependent variable (R&D expenditures net of public support)				
Explanatory variables:				
Regional subsidy	0.03 (1.88)*	0.07 (2.74)***	0.14 (2.22)**	0.05 (1.88)*
Research cooperation	0.03 (0.89)	0.14 (1.74)*	0.11 (1.51)	0.09 (1.44)
Young Innovative Company	-0.08 (-1.84)*	-0.02 (-0.34)	1.36 (12.64)***	-0.03 (-0.75)
PhDs and civil engineers	0.01 (0.23)	0.06 (1.78)*	0.03 (0.59)	0.11 (3.23)***
Master	0.06 (1.95)**	0.11 (2.51)***	0.11 (1.90)*	0.19 (4.09)***
Tax credit R&D	-0.03 (-1.71)*	-0.00 (-0.05)	-0.07 (-1.25)	0.33 (1.33)
Tax deduction 80% patent income	-0.02 (-1.06)	-0.03 (-1.06)	0.03 (0.20)	0.00 (0.07)
Control variables:				
Cash flow	0.18 (1.30)	0.09 (0.50)	0.06 (0.32)	0.20 (1.43)
Number of employees	0.97 (2.56)***	1.31 (2.58)***	0.07 (0.11)	1.01 (1.59)
Age	0.02 (0.09)	-0.00 (-0.02)	-0.02 (-0.07)	-0.20 (-0.68)
Capital intensity	0.23 (1.18)	0.48 (1.83)*	-0.15 (-0.27)	0.15 (0.61)
Region dummies	No	No	No	No
Industry (two-digit NACE)	No	No	No	No
Year dummies	Yes	Yes	Yes	Yes
Industry x year dummies	No	No	No	No
Number of observations	2,270	1,959	1,324	2,794

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) by Pavitt category, using the updated classification of industries provided by Boggliacino and Pianta (2015). All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

Table 20 Results of panel estimation by orientation of R&D activities (2003-2011)

	Basic research	Applied research	Experimental development
Dependent variable (R&D expenditures net of public support)			
Explanatory variables:			
Regional subsidy	0.06 (1.79)*	0.08 (3.31)***	0.07 (3.45)***
Research cooperation	0.07 (1.40)	0.05 (1.09)	0.14 (2.62)***
Young Innovative Company	-0.06 (-0.75)	0.03 (0.47)	-0.07 (-1.03)
PhDs and civil engineers	0.11 (2.40)**	0.04 (1.43)	0.08 (2.85)***
Master	0.15 (3.16)***	0.10 (3.24)***	0.09 (2.85)***
Tax credit R&D	0.14 (1.13)	-0.01 (-0.41)	-0.05 (-1.80)*
Tax deduction 80% patent income	0.06 (0.70)	-0.05 (-2.25)**	0.02 (0.48)
Control variables:			
Cash flow	0.34 (1.93)*	0.09 (0.77)	-0.01 (-0.06)
Number of employees	0.77 (1.38)	0.98 (2.28)**	1.33 (3.02)***
Age	-0.17 (-0.67)	-0.13 (-0.72)	0.01 (0.08)
Capital intensity	0.37 (1.52)	0.38 (1.78)*	-0.23 (-0.91)
Region dummies	No	No	No
Industry (two-digit NACE)	No	No	No
Year dummies	Yes	Yes	Yes
Industry x year dummies	No	No	No
R-squared (within)	0,05	0,08	0,03
Number of observations	2,314	3,339	3,236

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) by category of R&D orientation. Industries have been split into three groups based on their ranking in terms of the average share of R&D expenditures oriented towards basic research, applied research and experimental development. All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

The effect of subsidies is larger and more significant in industries that rely more on applied research or experimental development. On the other hand, the coefficients of the partial exemption for researchers with a PhD or civil engineer degree and researchers with a master degree are somewhat smaller in these industries. In industries with a relatively high share of applied research, the effect of the partial exemption for researchers with a PhD or civil engineer degree is not statistically significant. The effect of the partial exemption for R&D cooperation is only significantly positive in industries that tend to rely on experimental development. The most surprising result is the statistically negative coefficient of the tax deduction of 80% of patent income in applied research industries and of the tax credit in experimental development industries (at 10%).

3.2. Selection bias

As pointed out before, if the selection criterions of agencies that provide direct support and the self-selection and autonomy of firms to decide how much to invest in R&D and whether or not to apply for public support are not accounted for, regression may result in a biased estimate of the impact of public support.

Henningsen et al. (2015) considered the evaluation grades of applications by firms, for R&D subsidies granted by the Research Council of Norway. Evaluation grades appear to predict, to a large extent, R&D

investments of firms. However, the grades do not change much over time such that, at least in the sample of Henningsen et al. (2015), unobserved project quality can be absorbed by firm fixed effects. However, not all factors that can induce a selection bias can be expected to be fixed over time. A possible way to take into account the selection mechanism involved in public support for R&D is a so-called selection model, which consists in the estimation of two specifications, a selection model and the actual (structural) regression. The selection model attempts to asses which factors can explain why a firm receives support in a given year or not. The estimation of this model provides variables that can be included in the specification of interest, in effect the regression of private R&D expenditures on public support. The statistical significance of these variables will provide an indication on the relevance of the selection bias. The original selection model, proposed by Heckman (1979) considered a bivariate choice variable (for example, support or not), which would reflect the possible outcome for a single instrument. Busom (2000) and Hussinger (2008) have applied the two-step selection model in the context of evaluation of public support for R&D. In our case, in a given year a firm can receive direct support (subsidies) as well as one of six tax benefits. The selection and self-selection involved in regional subsidies and federal tax incentives is likely to be explained by different firm and industry characteristics. Rather than using a bivariate selection, four possible categories of public support are considered[28]:

– Firm receives no support for R&D
– Firm receives a subsidy but no tax benefit
– Firm receives a tax benefit but no subsidy
– Firm receives a subsidy as well as a tax benefit

More details on the estimation of selection models are provided in Annex 1.

In a review of studies that estimate the probability of receiving public support for R&D, da Silva (2014) enlists factors that are commonly considered. Firm size is often found to be a significant determinant. Usually, larger firms are more likely to receive a subsidy except for countries where agencies explicitly favour smaller firms. Past experience in R&D as well as past participations in R&D programmes also raise the probability of receiving a subsidy. Other factors that have been considered are the patent stock of companies (number of patents raises probability of support) and firm age, although results for the latter are more ambiguous. Human capital, as for example measured by the share of qualified personnel, also seems to increase the probability of receiving support. Foreign ownership, on the other hand, is in some studies found to negatively affect the probability of support, which may be due to a preference of agencies for domestic firms or research centres of foreign companies being located abroad (da Silva, 2014: p. 10). Given data availability and the determinants listed by da Silva (2014), lagged values of R&D expenditures and public support are included in the first-step estimation of the selection equation. Cash flow is included as a proxy for the extent to which companies can finance R&D activities out of own funds. As the regional agencies that grant subsidies in Belgium have specific programmes for SMEs, a dummy variable denoting whether a company is a SME (less than 250 employees) or not is included as the exclusion criterion in the first-step selection specification, following Takalo et al. (2013) who use a

28 Busom and Fernández-Ribas (2007) consider the participation of Spanish firms in national and European R&D programmes. Probit and Multinomial Logit estimates indicate that different factors explain the participation in European and national programs.

SME dummy as exclusion variable in their estimation of the returns to R&D subsidies, provided by the Finnish Funding Agency for Technology and Innovation (TEKES). Lagged values of R&D and public support are also included in the selection model but not in the baseline specification used to estimate the impact of public support on R&D expenditures.[29] To allow for a potential non-linear effect of firm age, a squared term is also included in the specification.[30] The results of the multinomial logistic regression of the selection specification are reported in Table 21.

Table 21 Determinants of receiving public support for R&D (2003-2011)

	Subsidy – no tax benefit	Tax benefit – no subsidy	Subsidy and tax benefit
Explanatory variables:			
Lag R&D expenditures (net of support)	1.08 (4.61)***	1.17 (7.42)***	1.31 (5.71)***
Lag regional subsidy	1.28 (20.90)***	1.03 (1.38)	1.28 (13.46)***
Lag tax support	1.05 (1.10)	1.73 (22.98)***	1.76 (20.27)***
Cash flow	0.88 (-3.50)***	1.21 (4.45)***	1.12 (1.97)**
SME (0/1)	0.66 (-2.13)**	0.70 (-1.79)*	0.71 (-1.36)*
Age	0.99 (-1.21)	0.98 (-2.12)**	0.97 (-3.06)***
Age^2	1.00 (1.21)	1.00 (0.96)	1.00 (2.72)***
Capital intensity	1.21 (4.13)***	0.99 (-0.03)	0.97 (-0.41)
Mc Fadden pseudo R-squared: 0.51			
Number of observations: 5,952			

Note: The table shows the results of multinomial logistic regression. The dependent variable is a category variable reflecting four possible situations in terms of public support for R&D in a given year: 1 (firm receives no support for R&D); 2 (firm receives a subsidy but no tax benefit); 3 (firm receives a tax benefit but no subsidy) and 4 (firm receives a subsidy as well as a tax benefit). The table shows the results for the latter three categories relative to the benchmark group of no support. The coefficients denote the relative risk ratio which reflects the change in probability to belong to a group, relative to the benchmark group, for a unit change in the explanatory variable, with the other variables held constant. The SME dummy equals 1 for SMEs (employees<=250) and 0 for large firms. The estimation considers region, industry and year dummies (not reported).

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%.

The coefficients denote the relative risk ratio, with firms that receive no support in a given year taken as the benchmark group. Taking, for example, the coefficient for lagged R&D expenditures in the group of firms that receive a subsidy but no tax benefit, it implies that a positive unit change in lagged R&D expenditures increases the odds to receive a subsidy but no tax benefit by 1,08 relative to the odds to receive no support.

A coefficient above (below) one indicates – if the coefficient is statistically significant – that the higher the given variable the higher the odds to belong to a given group of public support relative to receiving no support.

The coefficient of lagged R&D expenditures is statistically significant and higher than 1 for the three groups of firms that receive some form of public support. This is in line with the importance of past R&D activities found in the literature. Past experience explains most the fact that firms receive direct support as well as some tax benefit and least, although still substantially so, whether firms only receive direct support (subsidies). Reflecting the evidence of persistence in public support, the fact that a firm

29 In some estimations to assess the robustness of the results of the benchmark specification, lagged variables are included. The one-year lagged total amount of the six different tax benefits is considered as the multinomial logistic regression, which is an iterative procedure, fails to converge if the tax benefits are broken down by scheme.

30 Estimations without a squared term for firm age provides similar results.

receives support in a given year is also significantly explained by the fact that the firm received support in the previous year.

The coefficient of cash flow is statistically significant for all three groups but for subsidies it is below 1 and for the two other groups above 1. This indicates that cash flow is an important factor for firms to apply for and receive tax support whereas cash-constrained firms are relatively more inclined to apply for direct support.

The coefficient of the SME variable indicates that SMEs are relatively less inclined to apply for public support. Somewhat surprisingly, this seems to be especially the case for direct support.

Firm age does not seem to play a significant role in receiving a subsidy. Relative to firms that receive no support, older firms – conditional on the other explanatory variables – are less likely to receive a tax benefit but no subsidy. For the group of firms that receive direct support as well as some tax incentive, the estimates of the linear and the squared age term are both statistically significant but with an opposed effect on the relative odds, suggesting that older firms have a higher probability, relative to firms without support, to belong to this group.

Finally, capital intensity is an important determinant of firms receiving a subsidy. Some of the industry, region and year dummies that are included in the selection model but not reported, are statistically significant. The Mc Fadden's pseudo R-squared of 0.51 indicates that the explanatory variables considered in the selection model explain a considerable part of the fact to which group of public support firms belong.

The results of the estimation of the selection model reveal that a number of firm characteristics can explain the fact that a firm will receive a given kind of support or combines support schemes. Although the tax benefits provided by the federal government are not subject to specific firm characteristics, except for the Young Innovative Companies scheme, firm characteristics such as past experience and support but also cash flow, firm size and firm age significantly explain whether firms apply for and receive tax benefits for their R&D activities or not.

To assess whether the (self-) selection affects the results of the actual regression, Heckman (1979) computes the inverse of Mill's ratio from the first-step selection model estimation. This ratio is introduced in the second step estimation. Statistical significance of the coefficient of the inverse Mills ratio indicates that ignoring (self-) selection of firms is likely to result in biased estimates. The original bivariate Heckman model provides one inverse Mills ratio. In a multivariate context, (*number of categories - 1*) inverse Mills ratios need to be computed (Dubin and McFadden 1984), so in the case of four categories three ratios need to be computed. Table 22 shows the results of the estimation of the baseline specification in which the three inverse Mills ratios, from the selection model estimation, are included.

Table 22 Results of panel estimation accounting for (self-) selection (2003-2011)

	Selection (incl. lag R&D expenditures)	Selection (no lag R&D expenditures)
Explanatory variables:		
Regional subsidy	0.02 (1.32)	0.05 (2.92)***
Research cooperation	0.02 (0.66)	0.05 (1.41)
Young Innovative Company	0.01 (0.21)	0.05 (0.71)
PhDs and civil engineers	0.01 (0.68)	0.02 (1.00)
Master	0.03 (1.32)	0.07 (2.78)***
Tax credit R&D	0.08 (2.89)***	0.02 (0.79)
Tax deduction - 80% patent income	0.02 (0.94)	0.01 (0.23)
Inverse Mills (subsidy – no tax benefit)	0.16 (7.91)***	0.03 (1.57)
Inverse Mills (tax benefit - no subsidy)	0.28 (8.40)***	0.05 (1.45)
Inverse Mills (subsidy and tax benefit)	0.27 (12.79)***	0.07 (2.78)***
Control variables:		
Cash flow	0.01 (0.14)	0.00 (0.04)
Number of employees	-0.06 (-0.17)	0.31 (0.74)
Age	0.60 (1.02)	0.25 (0.31)
Capital intensity	-0.02 (-0.12)	-0.00 (-0.02)
R-squared (within)	0.31	0.14
Number of observations	5,212	5,350

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification (1) in which three inverse Mills ratios, from the estimation of the selection model, reported in table 21, are included to account for (self-) selection. The last columns show the results of an estimation in which three inverse Mills ratios are included that have been derived from a selection model estimation in which lagged R&D expenditures are not considered as potential determinant. All variables are considered in logs, except firm age and the Mills variables. The estimations include year and industry x year dummies.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

The results differ dramatically from the results of the baseline specification without inverse Mills ratios. Only the positive coefficient of the tax credit for R&D investment is statistically significant. The fact that the coefficient of the three inverse Mills ratios are also highly significant suggests that the (self-) selection needs to be accounted for.

The results in the estimation of the baseline specification with inverse Mills ratios included is predominantly driven by the inclusion of lagged R&D expenditures in the selection model. The last column in Table 22 shows the results of an estimation with three inverse Mills ratios from a first-step selection model estimation in which lagged R&D is not considered as a potential determinant.[31] These results are more in line with the baseline specification, providing a positive and statistically significant positive estimate for subsidies and the partial exemption for researchers with a master degree.

On the other hand, only the coefficient of the inverse Mills ratio of the group of firms that receive direct support as well as some tax benefit is statistically significant. The R-squared of this model is substantially lower than in the model based on a selection model including lagged R&D expenditures, indicating that the fit of the latter model is better.

As mentioned before, the inclusion of a lagged dependent variable, which in the two-step model occurs indirectly, is the topic of debate with scholars who point out the problem of a lagged dependent variable

31 The results of this estimation are not reported but the results are in line with those reported for the selection model with lagged R&D included.

in the presence of serial correlation or non-stationary variables (Achen 2001; Keele and Kelly 2006) as opposed to scholars who argue that ignoring the lagged dependent is likely to result in biased estimates (Wilkins 2014).

3.3. Instrumental variable estimation

A regression provides an indication of correlation between the left-hand-side ("dependent") variable and the right-hand-side variable(s) and no proof of any causal link. A problem that often arises when using economic variables is endogeneity of right-hand-side variables. An explanatory variable is said to be endogenous if it is correlated with the residual term. This violates a crucial assumption of OLS and panel estimations so that the consistency[32] of the estimates is no longer guaranteed. The potential endogeneity in the baseline specification is closely related to the issue of self-selection. Self-selection reflects the fact that whether a firm receives support or not, always starts with the decision of the firm to apply for support. In the baseline specification, which regresses the amount of the own R&D expenditures of a firm on the amount of support it receives for its R&D activities, endogeneity of the support variables follows from the fact that it is the firm which decides to apply for which kind of support and, especially for the tax benefits, for which amount. David et al. (2000) argued that early studies on public support for R&D, by ignoring endogeneity, probably resulted in biased (optimistic) estimates of input additionality. The most common econometric approach to account for endogeneity is instrumental variable (IV) estimation. The idea is to consider variables that are exogenous, which in contrast with endogenous variables are not correlated with the residual term, but are correlated with the endogenous variable, the "causal" effect of which we are interested in. In most cases valid instruments are however very difficult to find. Blundell and Costa Dias (2009) argue that in models with unobserved heterogeneity across individuals, an IV estimator will only provide an average treatment effect under strong assumptions that are not likely to hold in practice. More details and a discussion of the advantages and limitations on IV estimation are provided in Annex 1.

In the context of the evaluation of public support for R&D, a good instrument is a variable that reflects a decision of a government, or the agency that provides the support, that is likely to affect the decision of a firm to apply for support but does not affect the private R&D expenditures of the firm. Lichtenberg (1988) argues that the total value of contracts that are awardable to a firm in a given year can be considered as an instrument for the support received by that firm. He uses the total value of contracts received by all firms with the same two-digit product and services code that the firm has sold to the government. Wallsten (2000) uses a similar instrument and Clausen (2008) considers the total amount of funding at the industry level as an instrument for public support received by firms.

For subsidies two instruments are considered: the average subsidization rate (total amount of subsidies received over total R&D expenditures) in the industry in which the firm is classified and the total amount of subsidies (net of subsidies received by the firm) in the firm's industry. For both instruments industry is defined at the three-digit NACE level. The instruments for the tax incentives that are con-

[32] A sample estimator is consistent if it approaches the true population value as the sample reaches infinity. An estimator of a parameter is unbiased if its expected value (mean) equals the true value. An estimator is efficient if it has minimum variance.

sidered follow the recent studies by Chang (2012) and Rao (2013) who use changes in tax policy to construct instruments for the public support variables. The instruments used for subsidies and the six tax support schemes are detailed in Table 23.

In addition to the instruments listed in Table 23, lags of all support variables are also included. This permits to test whether the instruments that are used are valid and relevant. Although an IV estimator is in principle more consistent than OLS, if some of the explanatory variables are endogenous, IV estimation can result in substantial efficiency loss (e.g., Cameron and Trivedi 2010), in effect, underestimation of the statistical significance of estimates. Good instruments need to satisfy two conditions: they should not be correlated with the residual term of the actual regression (valid instrument) but they should be strongly correlated with the endogenous variable they "instrument" (relevant instrument). Tests of these conditions require that the model is over-identified, in effect, that there are more instruments than endogenous variables.

IV estimations is traditionally performed in two steps (Two-stage least squares estimation). In a first step the potentially endogenous variable is regressed on its instrument(s). The prediction of the "endogenous" variable that results from this estimation is then used as a right-hand-side variable in the second step estimation. Under the assumption that the instruments are valid this estimation will provide a consistent estimate of the impact of the endogenous variable on the dependent variable of the second step regression. As in the baseline specification, all seven variables of public support are potentially endogenous, a two-stage least squares estimation implies seven first-step estimations, in which each support variable is regressed on the instruments and control variables. The results of the seven first-step regressions are not reported but the Angrist-Pishke under-identification tests reported in Table 24 provide an indication of the relevance of the instruments. Rejection of the null hypothesis of the test indicates that the coefficients of the instruments in the first-step regressions are statistically significant.

Table 23 List of instruments

	Instrument 1	Instrument 2
Regional subsidy	Average subsidization rate by three-digit industry	Total amount of support (net of support firm) by three-digit industry
Research cooperation	Share of firms that cooperate in three-digit industry * (rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year)	Rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year
Young Innovative Company	Share of young firms (age<=10 years) in three-digit industry * (rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year)	Rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year
PhDs and civil engineers	Average share of researchers with a PhD in total number of employees by three-digit industry * (rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year)	Rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year
Master	Average share of researchers with a university degree in total number of employees by three-digit industry * (rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year)	Rate of partial exemption if firm did not receive same benefit in previous year OR change in rate of partial exemption if firm also received this benefit in previous year
Tax credit R&D	Total amount of tax credit by three-digit industry	Applicable rate of deduction if firm did not receive a tax credit in previous year OR change in the rate of deduction if firm also received tax credit in previous year
Tax deduction 80% patent income	Total amount of tax deduction by three-digit industry	Applicable rate of deduction if firm did not receive a tax deduction in previous year OR change in the rate of deduction if firm also received tax deduction in previous year

The predictions of the support variables derived from these seven first-step estimations are then used as right-hand-side variables in the actual regression. More detail on the estimation and the tests are provided in Annex 1.The results of the second step are reported in Table 24.[33]

[33] IV estimations have been done using the Stata procedure XTIVREG2.

Table 24 Results of instrumental variable estimation (2003-2011)

	Instruments 1	Instruments 2	Combined instruments (Best first stage results)	Random Effects
Explanatory variables:				
Regional subsidy	0.12 (1.30)	0.45 (2.43)**	0.07 (0.59)	0.26 (6.51)***
Research cooperation	0.06 (0.88)	0.03 (0.34)	0.08 (1.01)	0.07 (1.42)
Young Innovative Company	-0.08 (-0.50)	-0.00 (-0.01)	0.02 (0.09)	0.19 (3.28)***
PhDs and civil engineers	-0.03 (-0.78)	0.02 (0.51)	0.03 (0.65)	0.14 (4.81)***
Master	0.13 (3.10)***	0.06 (1.02)	0.11 (2.22)**	0.10 (3.09)***
Tax credit R&D	0.00 (0.00)	0.01 (0.19)	-0.02 (-0.29)	0.01 (0.19)
Tax deduction 80% patent income	-0.02 (-0.29)	0.00 (0.03)	-0.02 (-0.23)	-0.00 (-0.00)
Control variables:				
Cash flow	0.06 (0.64)	0.01 (0.10)	0.08 (0.80)	0.24 (4.01)***
Number of employees	0.93 (3.15)***	0.66 (1.95)**	0.86 (2.69)***	0.33 (4.16)***
Age	0.32 (0.54)	-0.37 (-0.53)	0.22 (0.33)	-0.39 (-3.41)***
Capital intensity	0.08 (0.61)	0.03 (0.20)	0.03 (0.24)	0.00 (0.03)
Sargan (over-identification)	13.93 (0.05)**	5.54 (0.59)	9.28 (0.16)	
Anderson (under-identification)	71.83 (0.00)***	44.42 (0.00)***	98.07 (0.00)***	
Weak instrument(robust):				
Anderson-Rubin F	2.15 (0.00)***	1.76 (0.04)**	1.69 (0.06)*	
Anderson-Rubin Chi2	33.49 (0.00)***	26.92 (0.02)**	23.83 (0.03)**	
Stock-Wright	33.04 (0.00)***	26.72 (0.02)**	23.67 (0.03)**	
Angrist-Pishke (under-identification):				
Regional subsidy	75.47 (0.00)***	45.42 (0.00)***	102.48 (0.00)***	
Research cooperation	677.11 (0.00)***	835.67 (0.00)***	806.91 (0.00)***	
Young Innovative Company	799.40 (0.00)***	1031.25 (0.00)***	1030.47 (0.00)***	
PhDs and civil engineers	1080.62 (0.00)***	1590.29 (0.00)***	1532.65 (0.00)***	
Master	1141.02 (0.00)***	1288.57 (0.00)***	1336.37 (0.00)***	
Tax credit R&D	1481.26 (0.00)***	1782.93 (0.00)***	1735.21 (0.00)***	
Tax deduction - 80% patent income	839.35 (0.00)***	1103.94 (0.00)***	1039.34 (0.00)***	
R-squared	0.17	0.05	0.12	0.07
Number of observations	3,451	4,905	4,654	5,682

Note: The table shows the results of the second step of an instrumental variables estimation. The second column shows the results of an estimation in which instruments 1, listed in Table 23 are used in addition to lags of the support variables. The third column shows the results of an estimation in which instruments 2 are used in addition to lags of the support variables. The fourth column shows the results in which the best instrument for each support variable, based on the first stage results, is considered. The last column shows the results of an IV estimation with random effects.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

The second column shows the results of an estimation in which the instruments 1 listed in Table 23 are considered in addition to the lags of the support variables whereas the third column shows the results of an estimation in which instruments 2 listed in Table 23 are considered in addition to the lags of the support variables.

The results of the tests of the validity and relevance of the instruments are also reported. The Sargan test of over-identifying restrictions tests the crucial assumption that the instruments are not correlated with the residual term of the actual (second step) regression. As the null hypothesis is zero correlation, rejection provides an indication that (some) instruments are not valid or more generally that the model is not correctly specified. The relevance of instruments is tested with under-identification or weak instruments tests. The null hypothesis is zero correlation between the instrument(s) and the instrumented (endogenous) variable so that failure to reject would indicate that the instruments are not relevant or weak. Both under-identification and weak identification tests for all instruments and tests for each instrumented support variable separately are reported.

The loss in efficiency of IV estimation is reflected in the fact that for both lists of instruments, the effect of only one support variables is statistically significant. Using the first list of instruments provides a significant positive coefficient for the partial exemption for researchers with a master degree whereas for the estimation with the second list of instruments the positive – and rather high – coefficient of subsidies is significant. For the first list of instruments all under-identification and weak instruments tests are clearly rejected, indicating that instruments appear to be relevant but the null hypothesis of the Sargan test is also rejected, which casts some doubt on the validity (exogeneity) of the instruments. In the estimation with the second list of instruments, the null hypothesis of the Sargan test is not rejected whereas the under-identification and weak instrument tests are rejected although the Anderson-Rubin and the Stock-Wright test provide some indications, though only at the 1% significance level, that instruments may be weak. The R-squared of the latter estimation is also substantially lower than in the estimation with the first list of instruments, indicating a lower goodness of fit.

The fourth column in Table 24 shows the results of an IV estimation in which the instruments from Table 23 that perform best in the first stage result (correlate best with the variable they instrument) are considered. For regional subsidies, the average industry-level subsidization rate performs better than the total amount of subsidies received by firms in the same industry (NACE three-digit). The latter instrument is actually not statistically significant in the first stage regression of regional subsidies. For the tax support variables, the second instrument performs better than the first instrument. For the partial exemption for researchers with a master degree, none of the two instruments performs well and in the specification reported in the last column, only the lagged value is considered as an instrument for this support variable.

As the lags generally perform well in the first stage regressions they are all included in this specification. The specification has a better goodness of fit than the specification with the second instruments and in contrast with the specification with the first instruments, the null hypothesis of the Sargan test is not rejected, suggesting that the instruments are valid. There are however even more indications that some of the instruments are weak as the Anderson-Rubin Wald F-test is only rejected at the 10% significance level.

Lichtenberg (1988) points out that IV estimation with fixed effects will only provide good results if instruments are endogenous with respect to omitted time-invariant characteristics. The last column in Table 24 therefore also shows the results of a random effects (Generalized Least Squares) estimation. The results of this specification are more in line with the results of the baseline specification, with statistically significant positive – and rather large – effects for subsidies and three schemes of partial exemption for researchers.

The irony of instrumental variables estimation is that a good instrument for public support of R&D requires some complexity (threshold and conditions) and changes in government support whereas simple rules and stability of government support are generally recommended.

3.4. Matching

Matching is an alternative approach to the selection model and IV estimation that is often used to tackle the selection and endogeneity problem in the evaluation of public support. The rationale for matching is to approximate the ideal of randomized trial as much as possible with observational data. For each individual (firm) that receives a given treatment (e.g. R&D subsidy) another individual is considered that matches that individual as much as possible except for the fact that it does not receive the treatment. The outcome of the group of firms that receive support (treatment group) can be compared with the group of matched firms that do not receive support (control group). If all relevant characteristics that affect the probability that a firm receives treatment are considered in the matching process, the difference in the outcome between the treatment group and the control group provides an indication of the causal effect of the treatment. More details on matching are provided in Annex 1.

Almus and Czarnitzki (2003) was one of the first studies in which matching was applied to assess the effects of public support for R&D. Using data on R&D subsidies received by firms located in Eastern Germany, they find evidence of a statistically significant impact on R&D intensity. Czarnitzki and Lopes-Bento (2013) apply matching to R&D grants provided by the Flemish funding agency (IWT) and find indications of input additionality, which appears to be larger for SMEs than for large firms.[34] The authors consider the main advantage of matching over other (parametric) procedures to be that it does not rely on assumptions on the functional specification or the distribution of the residuals. Czarnitzki and Lopes-Bento (2014) apply matching in the evaluation of European and national innovation subsidies. The different sources of funding appear to be complements for German firms and public support also have a positive impact on innovation (more valuable patents).

Different matching procedures have been developed. Most procedures use the propensity score, in effect, the probability that a firm will receive support – conditional on the characteristics (covariates) that are considered – to match a firm that receives support with a firm that does not receive support. It is beyond the scope of this paper to compare these procedures (see for a recent review Caliendo and Kopeinig 2008). In this paper, the most straightforward matching procedure, nearest neighbour matching, is considered, which matches a firm with support with a firm without support that is closest in terms of the propensity score. The procedure allows for replacement so that a firm without support can be used to be matched with several firms that receive support. Matching with replacement increases the quality of matches and reduces the bias (Caliendo and Kopeinig 2008: p. 41). Following, for example, Czarnitzki and Lopes-Bento (2013) and Hottenrott and Lopes-Bento (2014), a caliper is applied to avoid bad matches that exceed a given maximum distance between the treated and the non-treated firm. The results of the matching procedure are reported in Table 25. Given that panel rather than cross-section data are used, a perfect match on the year is imposed.

[34] Similar results are found for the effect of regional subsidies in the estimation by firm size, as reported in Table 17.

Table 25 Average treatment effect on the treated – Nearest neighbour matching (2003-2011)

	Net R&D expenditures	Net R&D intensity
No matching on lagged R&D:		
Regional subsidy	0.62 (7.63)***	2.75 (5.47)***
Research cooperation	1.06 (4.25)***	11.00 (5.23)***
Young Innovative Company	0.67 (1.05)	16.00 (2.26)**
PhDs and civil engineers	1.98 (9.15)***	9.40 (7.60)***
Master	0.90 (4.01)***	1.10 (0.76)
Tax credit R&D	1.43 (3.07)***	6.31 (2.63)***
Tax deduction 80% patent income	0.00 (0.00)	-3.05 (-1.53)
Matching on lagged R&D:		
Regional subsidy	0.34 (5.09)***	1.83 (5.76)***
Research cooperation	0.80 (4.83)***	7.57 (4.34)***
Young Innovative Company	0.84 (3.02)***	14.58 (1.97)**
PhDs and civil engineers	1.98 (12.15)***	8.51 (9.28)***
Master	0.84 (5.24)***	2.89 (2.60)***
Tax credit R&D	1.36 (3.37)***	6.58 (2.70)***
Tax deduction 80% patent income	-0.01 (-0.01)	-2.93 (-1.14)

Note: The table shows the average treatment effect on the treated (ATET) resulting of a nearest neighbour matching procedure with caliper. Lagged values of support and control variables are used as pre-treatment covariates. The year is considered for a perfect match to avoid that firms are considered as their own match over time.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%.

Two alternative sets of covariates are considered to match firms. In addition to the four control variables that are used in all estimations (cash flow, number of employees, firm age and capital intensity) lags of the support variables are included. As pointed out in the discussion of the selection model, the fact that a firm receives support in a given year is often found to depend on whether the firm already received support in the past or not (see review by da Silva (2014) and the results in Table 21). The second set of covariates adds a more disputable variable, in effect, a lagged of the outcome variable (R&D input). This variable is not used in most studies. This can probably be explained by the fact that most matching studies are based on cross-section rather than on panel data although it may also be due to the well-known problems of including a lag of the outcome variable. On the other hand, given the reported strong persistence in R&D activities, the lag may reflect unobserved heterogeneity, which can bias the results of matching. The results of matching, respectively without and with a lag of R&D input are reported.

In many matching studies R&D intensity rather than the level of R&D expenditures are reported. For reasons of comparability, with other studies as well as with the other estimations in this paper, both alternative measures of R&D input are considered.[35] Table 25 shows the average treatment effect on the treated (see Annex 1 for more details). The results confirm the results of the baseline specification although the impact of the partial exemption for Young Innovative Companies and the tax credit for R&D investment is also found to be statistically significant. As in the baseline estimation, no significant effect can be detected for the tax deduction of 80% of gross patent income.

Matching considers the impact of treatment in a binary way, a firm receives support or not. Information on the amount of support received by firms is not considered. Recently, a number of studies have used

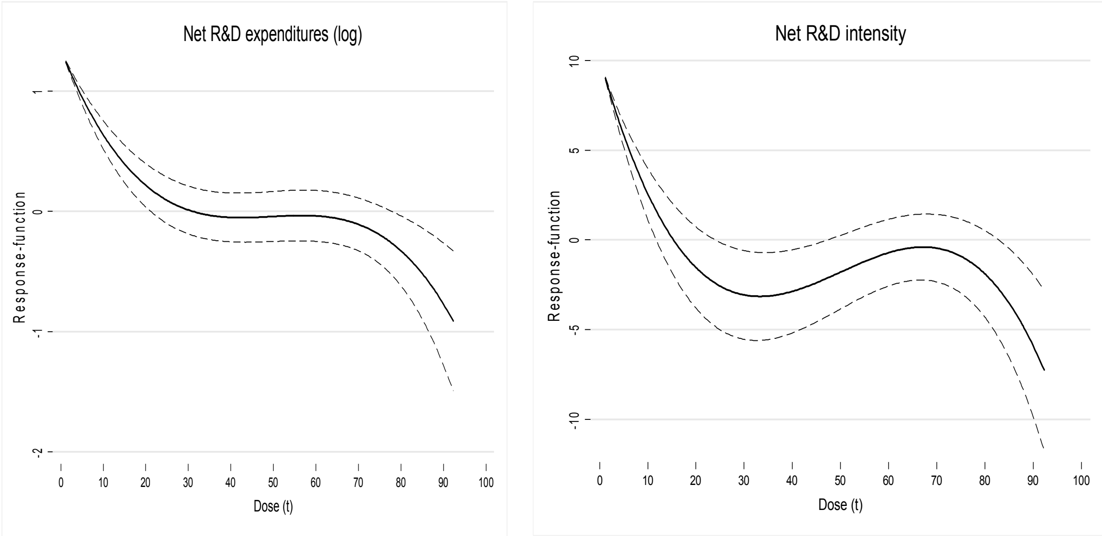
[35] Estimations were performed with the Stata procedure TEFFECTS (NNMATCH option).

an extension of matching procedures that permit to account for the amount of support. Marino et al. (2010) use a generalization of the propensity score that permits to consider the level of treatment. Their results suggest that for public funding of private R&D of Danish firms, substitution occurs for high levels of support. Hottenrott et al. (2014) use a generalization of the propensity score to estimate dose response functions to assess whether the size of R&D grants provided by the Flemish funding agency IWT matter for private R&D expenditures. Their results suggest increasing input additionality above a certain minimum subsidy threshold and indications that additionality drops again for large subsidies.

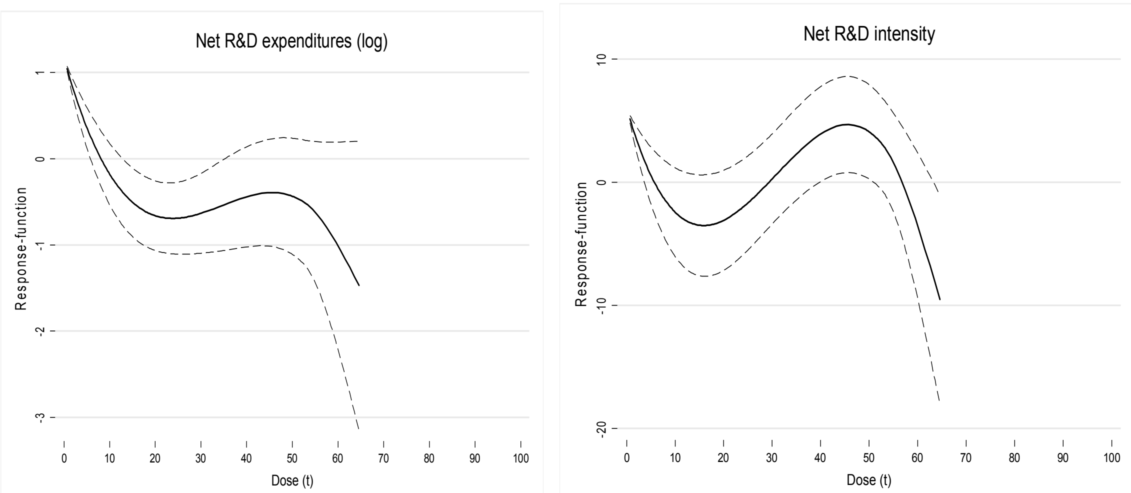
Running the dose response function proposed by Bia and Mattei (2008) provides indications of misspecification (e.g. violation of normality assumption). The Stata procedure CTREATREG proposed by Cerulli (2014) permits the estimation of dose response functions without the assumption of full normality. The procedure moreover permits many individuals to receive no treatment, which is the case in the dataset used in this paper as firms that receive no support outweigh firms that receive some support for R&D in any given year.

Cerulli and Poti (2014) applied the procedure to public support for R&D received by Italian firms over the period 1998-2006. They find indications of input additionality when net R&D expenditures is used as dependent variable but not when R&D intensity is the outcome variable. The results of the dose response functions are rather similar to the results of Hottenrott et al. (2014). The relationship between public support and R&D investment in Italy is positive but decreasing for a relatively low level of support and the effect appears to be negative for high doses, which the authors explain by the marginal increase of adjustment costs of higher R&D investment that are not fully compensated by public support. Graph 10 up to 16 show the results of the continuous treatment procedure by Cerulli (2014) for subsidies and the six tax incentives. The graphs show the impact of public support by increasing dose (ratio of support to R&D expenditures) on the x-axis, on respectively net R&D expenditures (log) and net R&D intensity. Especially for the schemes with relatively few benefiting firms the 5% confidence interval is very large and precludes any meaningful conclusion. For all schemes, the effect of support decreases with the rate of subsidization, though not always monotonically.

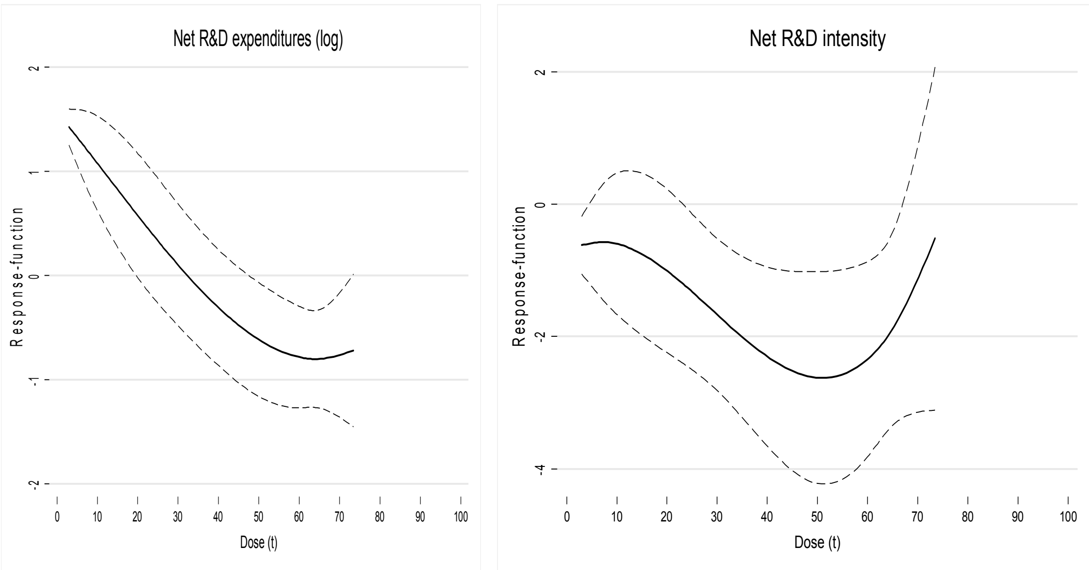
Graph 10 Dose response function regional subsidies



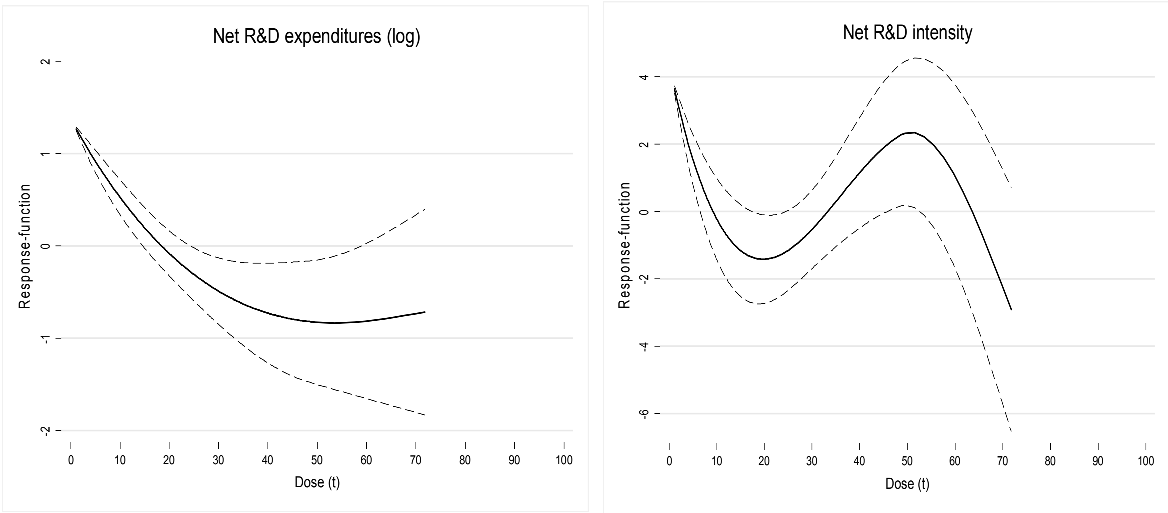
Graph 11 Dose response function partial exemption R&D cooperation



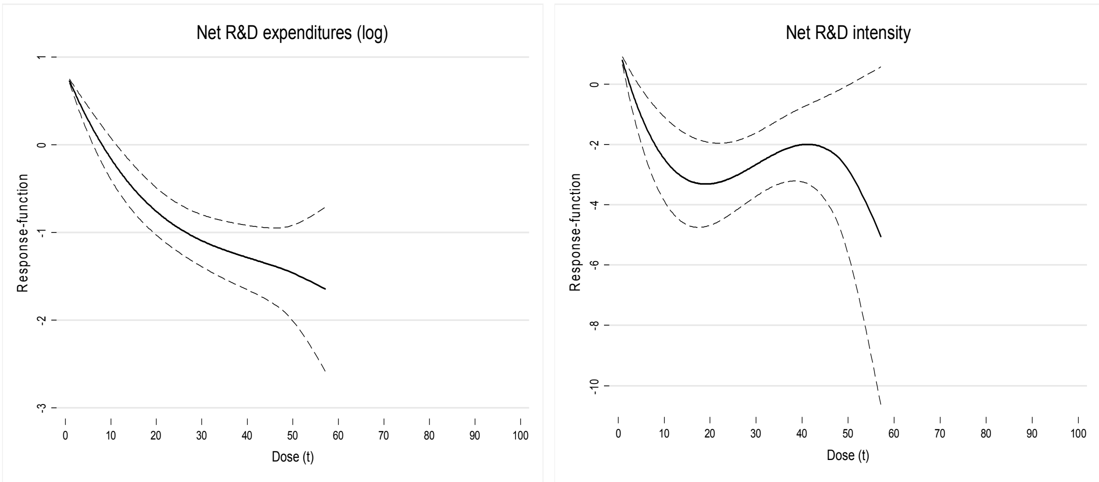
Graph 12 Dose response function Young Innovative Companies



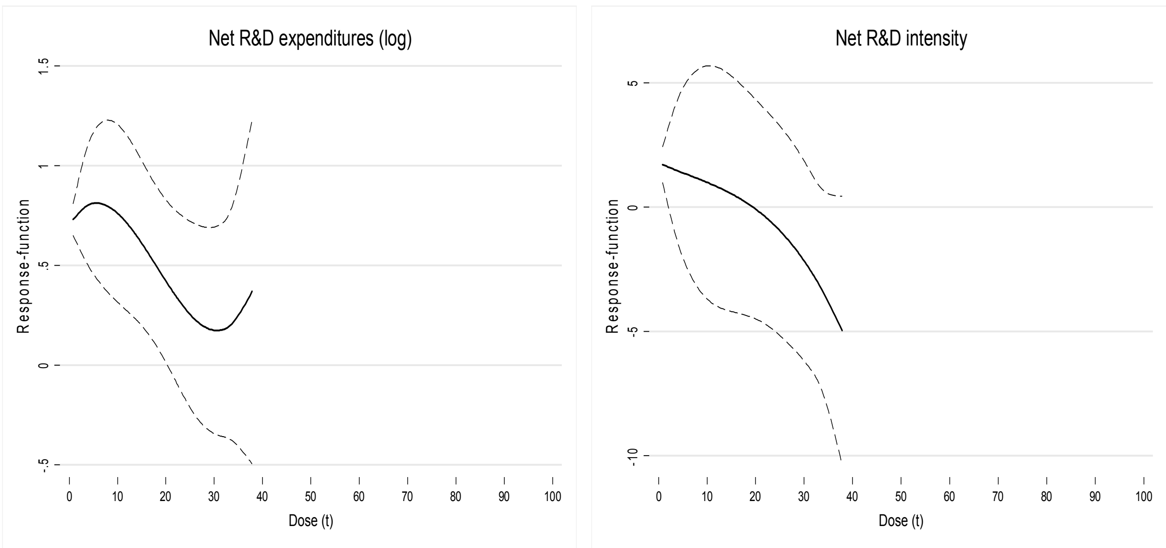
Graph 13 Dose response function partial exemption researchers with a PhD or civil engineer degree



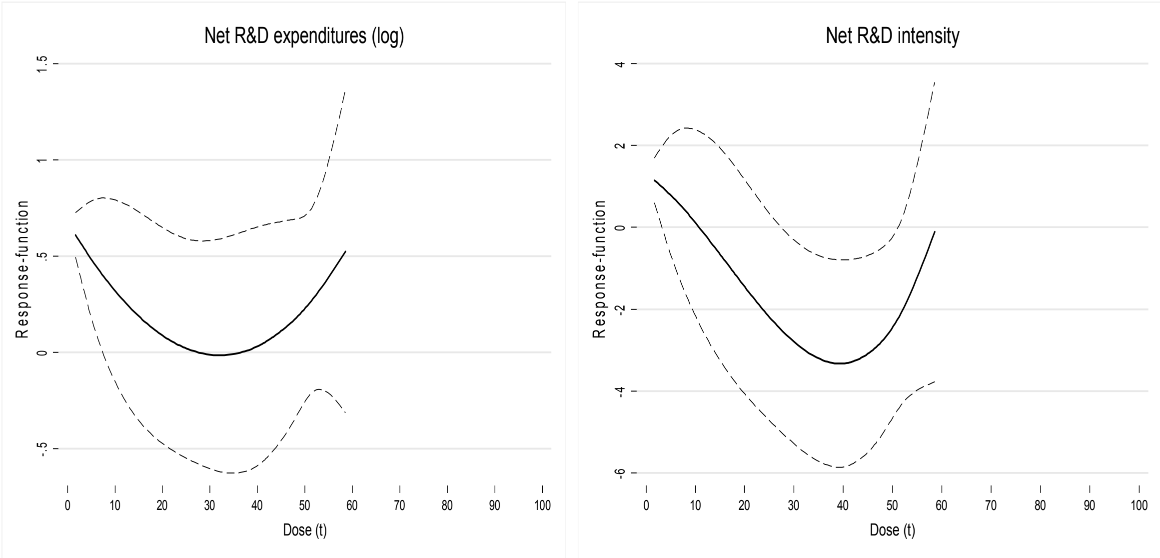
Graph 14 Dose response function partial exemption researchers with a Master degree



Graph 15 Dose response function Tax credit R&D investment



Graph 16 Dose response function Tax deduction 80% patent income



The results seem to be in line with the quantile graphs (graph 6 up to 9) and indicate that input additionality of most support schemes generally appears to decrease for increasing rates of support although there are also some indications of a threshold level of minimum support.

The continuous treatment regression used to create the graphs of dose response fits in a traditional OLS framework. The graphs do not result from matching on a generalized propensity score as in Marino et al. (2010) and Hottenrott et al. (2014) and are closer – though not easily comparable[36] – to the quantile graphs presented in section 3.5. A more detailed analysis using matching and dose response functions is beyond the scope of this paper but may be considered in future evaluations.

3.5. Dynamic panel

In section 3.2 the strong persistence in the R&D expenditures of firms was pointed out. When a lagged dependent variable is included in the baseline specification, its coefficient is substantial and statistically significant. Panel specifications with a lagged dependent variable may provide substantially biased results. Nickell (1981) showed that a fixed effects (within) estimator of a specification that includes a lag of the dependent variable is inconsistent so that even with a large number of individuals in the panel, estimates are not reliable. Although the bias applies to the lagged dependent variable, the estimates for the other explanatory variables may also be biased if these variables are correlated with the lag of the dependent variable (Baum 2013).

To account for the possible bias, dynamic panel procedures can be applied. The most straightforward dynamic panel procedure consists in first-differencing, which removes all time-fixed effects. Lagged levels of the first-difference variables are used as instruments. A more sophisticated procedure is the system GMM, which in addition to first differences also considers levels, for which first differences are

[36] The graphs can however not be compared in a straightforward way. The quantile graphs show how the estimates of the elasticity of R&D expenditures with regard to support vary with an increase in the amount of support whereas the dose response functions show how the effectiveness of support changes with an increase in the rate of support.

used as instruments. Both GMM procedures can be performed in one or two steps. Although the latter is asymptotically more efficient than the one-step estimation Arellano and Bond (1991) and Blundell and Bond (1998) pointed at a downward bias of standard errors in finite samples. Windmeijer (2005) proposed a finite-sample correction of the standard errors.

Table 26 Results of dynamic panel estimation (2003-2011)

	First-difference GMM (two-step)	System GMM (two-step)
Explanatory variables:		
Net R&D expenditures (t-1)	0.08 (2.28)**	0.84 (22.95)***
Net R&D expenditures (t-2)	-0.15 (-2.67)***	-0.15 (-2.79)***
Regional subsidy (t)	-0.02 (-1.54)	0.01 (0.68)
Regional subsidy (t-1)	0.00 (0.05)	0.04 (3.51)***
Research cooperation (t)	0.02 (0.62)	0.05 (2.77)***
Research cooperation (t-1)	-0.02 (-0.79)	-0.03 (-1.29)
Young Innovative Company (t)	-0.02 (-0.19)	0.01 (0.82)
Young Innovative Company (t-1)	0.09 (1.58)	0.10 (1.71)*
PhDs and civil engineers (t)	0.03 (2.29)**	0.09 (3.59)***
PhDs and civil engineers (t-1)	0.03 (2.34)**	-0.02 (-1.08)
Master (t)	0.01 (0.54)	0.03 (1.17)
Master (t-1)	0.06 (3.76)***	0.01 (0.69)
Tax credit R&D (t)	-0.01 (-0.49)	-0.02 (-1.13)
Tax credit R&D (t-1)	0.02 (1.22)	0.02 (0.93)
Tax deduction 80% patent income (t)	0.00 (0.13)	0.00 (0.15)
Tax deduction 80% patent income (t-1)	0.03 (1.41)	0.01 (0.57)
Control variables:		
Cash flow	0.04 (0.44)	0.11 (1.32)
Number of employees	0.61 (1.27)	0.24 (2.17)**
Age	-0.27 (-4.28)***	-0.01 (-1.43)
Capital intensity	-0.19 (-1.18)	-0.02 (-0.21)
Arellano-Bond test AR(1)	-2.20 (0.03)**	-7.18 (0.00)***
Arellano-Bond test AR(2)	-1.69 (0.09)*	-4.00 (0.00)***
Sargan (over-identification)	136.32 (0.00)***	544.24 (0.00)***
Hansen (over-identification)	82.12 (0.00)***	192.99 (0.00)***
Hansen test excluding group	4.35 (0.11)	100.54 (0.00)***
Difference (H_0=exogeneity)	77.77 (0.00)***	92.45 (0.00)***
Wald chi square (p-value)	63.83 (0.00)***	8779.53 (0.00)***
Number of observations	1,994	3,282
Number of instruments	50	59

Note: The table shows the results of a two-step first difference and a two-step system GMM with orthogonal deviations. For lags of the dependent variable GMM-style instruments are used and for the public support variables lags are used as instruments.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity and are, following Windmeijer (2005), corrected for the finite-sample bias of a two-step estimation.

In Table 26, the results of a two-step first difference and a two-step system GMM are reported. Given that the panel is highly unbalanced, forward orthogonal deviations are considered to avoid the loss of many observations due to the large number of gaps (More details on GMM are provided in Annex 1).

For both GMM procedures the Sargan as well as the Hansen test of over-identifying restrictions are strongly rejected, suggesting that the instruments are not valid. The Arellano-Bond tests also indicate second order serial correlation in the error term, indicating the specification is problematic. The results in Table 26 should therefore not be trusted too much. Mulkay and Mairesse (2013), in their evaluation of the French R&D tax credit also find that GMM provides poor results with rejection of the Sargan over-identification test and a problem of second order serial correlation.

The coefficients in the first-difference specification do not suggest a high degree of persistence in R&D expenditures, in contrast with the coefficients of the lagged dependent variable in the system GMM. Given the caveat of the obvious specification issues in both GMM specifications, the first difference GMM specification suggest a statistically positive effect of the partial exemption for researchers with a PhD or civil engineer degree and for researchers with a master degree. The System GMM specification, on the other hand provides statistically significant coefficients for regional subsidies, the partial exemption for Young Innovative Companies (10% significance level) and the partial exemption for researchers with a PhD or civil engineer degree. Alternative specifications (one-step, only one lag of the dependent variable or three lags, using GMM-style instruments or the instruments used in the IV estimation) provide similar results but all specifications generally perform poorly in terms of instrument validity and provide rather unreliable results.

A potential problem of a regression with variables considered over time is spurious correlation. Many economic variables, even expressed in constant prices, are not stationary but show a trend over time. If a trended variable is regressed on trended right-hand-side variables, the estimated coefficients which reflect the degree of association, are likely to be high even if there is no actual relationship between the variables, a problem that statisticians label as "spurious correlation". First differencing of trended often eliminates the non-stationarity of variables, at the cost of losing all level information. More interesting may be to establish whether there exists a real long-term relationship between non-stationary variables. An error-correction model attempts to distinguish the short-term effects from the long-run relationship between variables. Stationarity of time series is tested by unit root tests. Most unit root tests have a null hypothesis of non-stationarity so rejection indicates that the variable is not stationary. For two variables to be cointegrated – which implies a real long-term relationship – they have to share a common trend. An essential condition is that the variables have the same order of integration. A time series is said to be integrated of order 0 – I (0) - if the variable is stationary in levels. If the variable is not stationary in levels but its first difference is, the variable is integrated of order 1 – I (1). A variable is integrated of order 2 if it becomes stationary after second differencing.

A panel does not contain a single time series of a variable but a time series for each individual in the panel. The advantage of considering a panel is that unit root tests for single time series have low power.[37] Panel data may increase the power of unit root tests but a complication is the need to account for possible unobserved cross-section heterogeneity across individuals (Cameron and Trivedi 2010: p. 279). The panel unit root tests reported in Table 27 show that the variables of public support are highly non-stationary. Whereas private R&D expenditures appear to be integrated of order 1 (non-stationary in

[37] The power of a statistical test reflects the probability that a false null hypothesis is corrected rejected.

level but stationary in first differences), the time series of subsidies and the six tax incentives are still stationary after first differencing.

Table 27 panel unit root tests of levels and first differences (2003-2011)

	Level - test I(0)	First difference - test I(1)
Regional subsidy	857.36 (0.99)	1018.45 (0.00)***
Regional subsidy	1176.89 (0.33)	746.47 (1.00)
Research cooperation	565.56 (1.00)	483.81 (1.00)
Young Innovative Company	37.79 (1.00)	135.46 (1.00)
PhDs and civil engineers	1185.09 (0.27)	854.51 (0.99)
Master	154.30 (1.00)	231.17 (1.00)
Tax credit R&D	10.44 (1.00)	25.64 (1.00)
Tax deduction 80% patent income	11.10 (1.00)	22.98 (1.00)

Note: The table shows the results of a panel unit root tests on the level and first difference of R&D expenditures and the public support variables. The test has been performed using the Stata procedure XTFISHER which allows for unbalanced panels. The reported test is a Fisher panel augmented Dickey-Fuller unit root test. The null hypothesis is that the variable contains a unit root, in effect, is not stationary.

The *, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The reported values are the Fisher Chi-squares values with p-values in brackets.

From a cointegration perspective, these results are problematic as they suggest that R&D expenditures and public support cannot be cointegrated, in effect, that there is no long-term relationship between private R&D and public support. However, the concept of stationarity and (panel) unit root tests are only really meaningful for long panels, with observations over a substantially long period.

Keeping in mind the results of the panel unit root tests and the fact that the panel in this paper is rather short, Table 28 shows the results of an error-correction model (ECM) specification. This specification attempts to distinguish potential short –term effects of public support from the long-term (cointegration) relationship. More details on stationarity and error-correction models are provided in Annex 1. The error-correction coefficient, the coefficient of the lagged dependent variable in the ECM specification, is -0.65 which seems reasonable as this coefficient should be between -1 and 0. The long-term coefficients are in line with the results of the baseline specification, with a statistically significant positive coefficient for subsidies and the partial exemption for researchers with PhD or civil engineer degree at 10% significance level) and researchers with a master degree.

Although the results of the error correction model seem reasonable and in line with results of the baseline specification, the unit root test of the residual of the ECM indicates that the residuals are not stationary, which casts doubt on the existence of a cointegration relationship between R&D expenditures and public support, corroborating the different order of integration suggested by the panel unit root tests on the individual variables.

As mentioned before, stationarity, cointegration and error correction modelling are only meaningful for long time series (panels). This also applies to dynamic panel procedures.

Table 28 Results of an Error Correction Model (2003-2011)

	Short-term (First differences)		Long-term (lag levels)
Explanatory variables:			
R&D expenditures (t-1)			-0.65 (-26.41)***
Regional subsidy	0.02 (0.96)		0.05 (2.24)**
Research cooperation	0.04 (1.57)		0.02 (0.69)
Young Innovative Company	-0.05 (-0.71)		0.02 (0.33)
PhDs and civil engineers	0.03 (1.85)*		0.03 (1.76)*
Master	0.03 (1.37)		0.05 (2.16)**
Tax credit R&D	0.02 (0.92)		0.00 (0.05)
Tax deduction 80% patent income	0.01 (0.43)		-0.00 (-0.03)
Control variables:			
Cash flow		0.11 (1.14)	
Number of employees		0.88 (2.00)**	
Age		0.01 (0.08)	
Capital intensity		0.03 (0.18)	
Region dummies		No	
Industry (two-digit NACE)		No	
Year dummies		Yes	
Industry x year dummies		Yes	
R-squared (within)		0.41	
Number of observations		4,610	
Fisher panel unit root tests stationarity ECM residual:		0.00 (1.00)	

Note: The table shows the results of a one-step error correction model (ECM) specification. The second column shows the results of the short-term effects (coefficients of first differences of support variables) and the third column the coefficients of the long-term effect (coefficients of the lags of the support variables.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

As the panel used in this paper does not cover a very long period (especially concerning the public support variables), all results in this section should be interpreted with much caution. However, the strong persistence in both private R&D expenditures and public support also cautions against casual interpretation of estimations that do not take into consideration this persistence. Future evaluations, based on longer time periods should therefore elaborate on a more dynamic framework.

3.6. Starting R&D activities

Busom et al. (2015) point at the different policy perspective of raising the extensive margin of R&D activities, which consists in increasing the number of firms that perform R&D, and the intensive margin, which denotes the extent to which R&D active firms increase their R&D expenditures. They argue that high persistence of the status of firms (performing R&D or not), as suggested for Belgium by Of all firms in the dataset, for which data are available, 95.99% of firms with no R&D activities in a given year over the period 2003-2011 did not perform R&D in the next year and 92.25% of firms that did R&D in a given year also had R&D activities in the following year.

Table 9 hints at the failure to raise the extensive margin. It is not straightforward to model the decision of a firm to start R&D activities. Only if a firm responds to the R&D survey that it does R&D in a given year and has responded in previous years not to have performed R&D, can it be established that the firm started doing R&D. However, if a firm has not done any R&D in the past it is not likely to be on the list of R&D firms that is used to send the R&D survey. As such the number of firms that start doing R&D is likely to be underestimated. On the other hand, firms that only perform R&D occasionally may report no R&D for one or two years before reporting R&D again. These firms would not be real starters but rather re-starters. Given this caveat, Table 29 shows the results of a logistic regression in which the binary dependent variable reflects the decision of a firm to start doing R&D in a given year.

Table 29 Results of an estimation of the decision to start R&D activities (2003-2011)

	Fixed effects	Random effects
Explanatory variables:		
Regional subsidy	0.07 (0.66)	0.13 (3.96)***
Research cooperation	-	0.32 (4.20)***
Young Innovative Company	-	-0.08 (-0.47)
PhDs and civil engineers	2.03 (0.66)	-0.04 (-0.87)
Master	0.17 (2.08)**	0.04 (0.68)
Tax credit R&D	-	0.02 (0.17)
Tax deduction 80% patent income	-	-0.03 (-0.13)
Control variables:		
Cash flow	0.53 (1.06)	0.05 (0.59)
Number of employees	1.11 (0.76)	-0.34 (-3.13)***
Age	-0.07 (-0.00)	-0.02 (-1.18)
Age^2	0.00 (1.16)	0.00 (0.65)
Capital intensity	-1.19 (-1.62)*	-0.08 (-1.03)
Region dummies	No	Yes
Industry (two-digit NACE)	No	Yes
Year dummies	Yes	Yes
Industry x year dummies	No	No
LR (Wald for random effects) (prob> chi2)	189.98 (0.00)	225.18 (0.00)
Number of observations	378	1,044

Note: The table shows the results of a logistic regression of the decision to start R&D activities in a given year. The dependent variable is a dummy variable that equals 0 if a firm that did not perform any R&D in year t-1 also does not perform any R&D in year t and the dummy equals 1 if a firm that did not perform any R&D in year t-1 does report R&D in year t. Only real responses of firms as to whether or not they performed R&D are taken into consideration.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%.

The variable equals 0 for a firm that responds to the R&D survey that it has not performed R&D in year t as well as year t-1 and equals 1 if a firm responds that it performs R&D in year t but not in the previous year. As the estimation does not consider firms that perform R&D on a continuous base and only considers firms with real responses as to their R&D activities in two consecutive years, the number of observations is rather low. In a conditional fixed effects regression, shown in the second column, four of the support variables can actually not be estimated. Moreover the estimation fails to converge so the results are not reliable. The random effects estimation, reported in the third column, does converge, retains more observations and overall appears to provide more reliable results. Whereas the fixed effects estimation suggests that the partial exemption for researchers with a master degree has a significantly positive impact on the decision to start R&D, the random effects estimation suggests a statistically significant positive impact for regional subsidies and the partial exemption for firms involved in R&D cooperation.

3.7. Bang for the Buck (BFTB)

In many studies, input additionality of public support for R&D is expressed in terms of the Bang for the Buck (BFTB), which indicates how much one euro in public support fosters in additional R&D expenditures by companies. A linear regression of R&D expenditures on the amount of support provides a direct indication of the BFTB. Given the highly skewed distribution of R&D expenditures across firms a linear specification is likely to provide biased estimates of the effect of public support. In a log-linear specification the BFTB has to be computed based on the elasticity estimate and some benchmark value for the amount of support and the level of R&D expenditures.

Summarizing the main results in this chapter, Table 30 shows the Bang for the Buck, based on alternative estimates of the impact of subsidies and the tax incentives for R&D on private R&D expenditures. Only the statistically significant elasticity estimates are considered and the average of net R&D expenditures and the amount of support is considered as benchmark value. As R&D expenditures are considered net of the amount of public support, the BFTB provides a net effect. For example, the BFTB of 0.59 for regional subsidies in the fixed effects estimation indicates that one euro results in 1.59 euro R&D expenditures, 0.59 of which financed by private firms. The different estimation procedures provide fairly robust indications of a positive impact of direct support (subsidies) and the partial exemption for researchers with a master degree. The BFTB for these schemes is rather similar with IV estimation resulting in the highest and probably overoptimistic BFTB. For the partial exemption for researchers with a PhD or civil engineer degree, no statistically significant effects are found when selection and endogeneity are accounted for and the very high BFTB of the partial exemption for firms that are involved in R&D cooperation in the fixed effects specification is not confirmed by any of the alternative estimations. For the tax credit for R&D investment only the two-step selection model provides a (low) BFTB whereas for the tax deduction of 80% of gross patent income, no estimation provides any indication of a statistically significant impact on private R&D expenditures. Although the instrumental variables estimations perform rather well in terms of instrument validity and relevance, the very high estimates of the BFTB indicate that IV estimation is certainly not the panacea for the evaluation of public support.

Table 30 Bang for the Buck for significant elasticity estimates in alternative estimation procedures (2003-2011)

	Fixed effects	Selection	Instrumental variables	ECM (long term)
Regional subsidy	0.59	0.45 (without lag R&D)	3.94	0.43
Research cooperation	2.02-2.28	-	-	-
Young Innovative Company	-	-	-	-
PhDs and civil engineers	0.50-0.60	-	-	0.32
Master	1.37-1.51	1.03 (without lag R&D)	1.94	0.74
Tax credit R&D	-	0.23 (with lag R&D)	-	-
Tax deduction 80% patent income	-	-	-	-

Note: The table shows the Bang for the Buck (BFTB), an estimate of how much additional R&D expenditures result from one euro in public support received by companies. The BFTB is calculated at the mean of net R&D expenditures and support for non-missing observations in the given specification and only estimates of elasticity that are statistically significant (at least at 10%) are considered. For the fixed effects the range of the BFTB is shown for a specification with and without inclusion of industry x year dummies. For the results the BFTB is shown for the schemes for which the second-step coefficient is significant, based on a first-step selection specification with or without lagged R&D expenditures. For the instrumental variables estimation, the BFTB is shown for either the specification with the first or the second list of instruments. For the error-correction model (ECM) the BFTB is shown for those schemes for which the long-term elasticity is statistically significant.

4. Behavioural and output additionality

Chapter 3 considered the main research question in this paper, whether public support for R&D stimulates firms to spend more on their R&D activities than they would without support. Some recent studies focus on behavioural and output additionality of public support rather than on its input additionality. Behavioural additionality concerns effects on the characteristics of R&D activities such as the risk level, for example reflected in the share of R&D expenditures companies spend on basic research; applied research or experimental development. Empirical work on output additionality considers the effects of public support on outcome variables such as innovation (new products, processes or services), profits and productivity. In this chapter we evaluate the impact of subsidies and tax incentives in Belgium on the orientation of R&D activities of firms, on the composition of R&D personnel and on some output measures (productivity).

4.1. Orientation of R&D activities

Arora et al. (2015) posit that large companies appear to increasingly shy away from scientific research. The authors perceive a new division of labour in innovation activities with small start-ups that specialize in scientific research and larger firms that target product development and commercialization. Akcigit and Kerr (2010) report evidence for the US that small firms have a higher relative invention rate than large firms whereas large firms focus more on exploitation than exploration in R&D activities. Akcigit et al. (2014) advocate a policy of public support that targets basic research more directly than most subsidies or tax benefits, by increasing the intellectual property rights granted to academic researchers or by supporting collaboration between universities and the private sector (cf. the scheme of partial exemption for R&D collaboration in Belgium). Hottenrott and Lopes-Bento (2014) point at the evidence that young small firms tend to engage in more basic and radical innovation projects than large incumbents. In the matched data for Belgium, the correlation between the share of R&D expenditures that companies spend on basic research and firm size (number of employees) is negative but the correlation is not statistically significant. The correlation between the share of R&D expenditures that companies spend on applied research and firm size is -0.06 and statistically significant at the 5% level and the correlation between the share of R&D expenditures that companies spend on experimental development and firm size is 0.07 and also statistically significant at the 5%. Although these correlations are not very high they seem to corroborate the evidence for the US reported by Akcigit and Kerr (2010) and Arora et al. (2015).[38]

In this section, the potential impact of subsidies and the tax incentives on the share of R&D expenditures oriented towards different types of R&D activities is examined, considering the following question:

Do companies change the shares of their total R&D expenditures that they devote to respectively basic research, applied research and experimental development as a result of the subsidies or tax benefits they receive?

[38] The correlation between the three shares and firm size denoted by value added has the same sign as with the number of employees for firm size but none of the correlations is significant at 5%.

Table 31 shows the results of an estimation of the effects of subsidies and tax benefits on the orientation of R&D activities. In the R&D Survey, companies are asked to provide the distribution of their R&D expenditures over three categories: basic research, applied research and experimental development. The response to this question only applies to one of the two years that the biennial survey covers. For the estimation of the impact on the orientation of R&D activities, a pseudo panel is created, dropping the years for which there is no information. The pseudo panel considers the data for a given year (t) and the data two years before, which is considered as the previous period (t-1). This panel obviously contains fewer observations than the full panel.

Table 31 Results of panel estimation by orientation of R&D activities (2003-2011)

Dependent variable:	Basic research	Applied research	Experimental development
Explanatory variables:			
Regional subsidy	0.001 (3.33)***	0.001 (1.05)	-0.003 (-2.40)**
Research cooperation	0.001 (0.48)	-0.000 (-0.11)	-0.000 (-0.07)
Young Innovative Company	-0.001 (-0.37)	0.011 (2.86)***	-0.010 (-2.67)***
PhDs and civil engineers	-0.000 (-0.08)	0.000 (0.12)	-0.000 (-0.08)
Master	-0.001 (-1.13)	0.001 (0.61)	-0.003 (-0.15)
Tax credit R&D	-0.000 (-0.14)	-0.005 (-1.57)	0.005 (1.62)*
Tax deduction 80% patent income	0.003 (2.50)**	0.002 (0.48)	-0.005 (-1.50)
Control variables:			
Cash flow	0.002 (0.80)	-0.002 (-0.32)	-0.000 (-0.05)
Number of employees	-0.006 (-1.88)*	-0.012 (-1.54)	0.02 (2.31)**
Age	-0.000 (-1.88)*	-0.000 (-0.09)	0.000 (0.78)
Capital intensity	0.002 (1.11)	-0.004 (-0.70)	0.001 (0.27)
Region dummies		Yes	
Industry (two-digit NACE)		Yes	
Year dummies		Yes	
Industry x year dummies		No	
Number of observations		3,050	

Note: The table shows the results of a seemingly unrelated regression, using the shares of R&D expenditures that firms spend on respectively basic research; applied research and experimental development.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%.

As the specification considers different shares, with the same explanatory variables, Seemingly Unrelated Regression (SUR) is performed (e.g. Greene 2000) using the share of R&D expenditures that firms spend on respectively basic research; applied research and experimental development.[39] The results suggest that the direct support provided by the regional agencies motivates firms to shift away from experimental development towards more basic research. Although there are few indications of input additionality for the partial exemption for Young Innovative Companies, this scheme seems to induce these companies to shift efforts from experimental development to applied research. The tax deduction of 80% of patent income – for which no evidence of input additionality is found in Chapter 3 – appears to have a statistically significant positive impact on the share of R&D expenditures that firms spend on

39 As the shares sum to one, one of the share equations needs to be dropped. The STATA procedure SUREG permits to consider all equations at once. Estimations with one of the three shares dropped sequentially provide similar results.

basic research.[40] The negative coefficient for basic research (significant at 10%). The statistically significant positive coefficient for experimental development of the firm size variable (number of employees) and the negative coefficient (significant at 10%) of firm age for basic research again seem to support the view that large and older firms focus more on development and less on basic research.

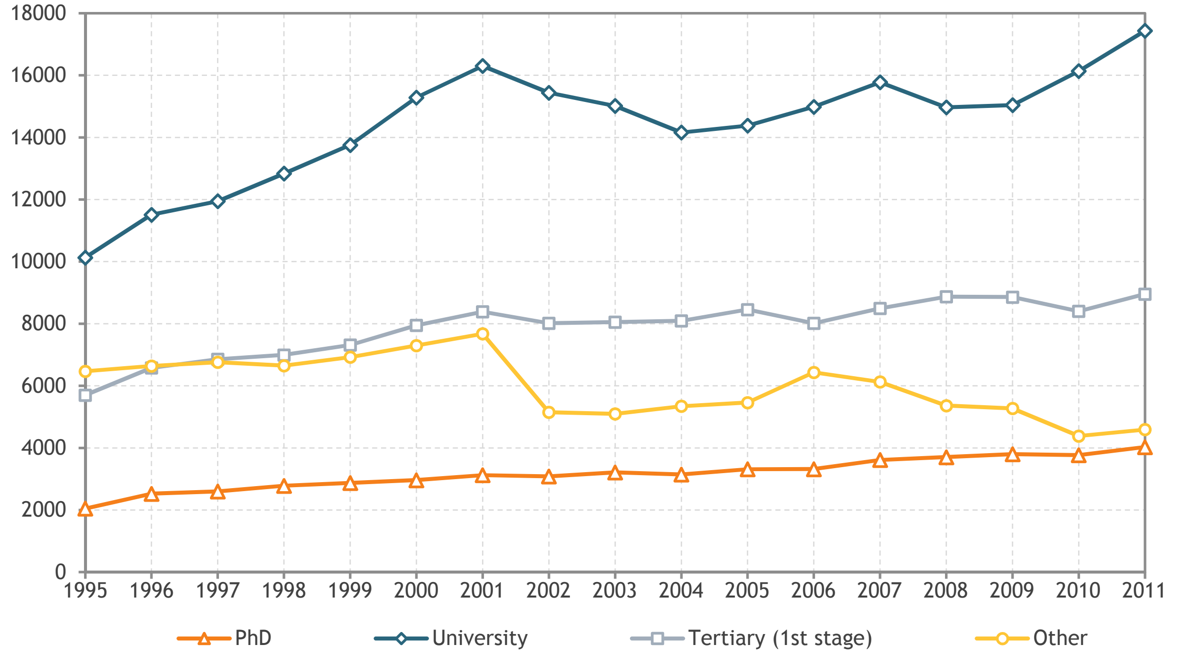
A study by Hottenrott et al. (2014) on support provided by the Flemish funding agency (IWT) that distinguishes between grants for research and grants for development activities, finds positive direct effects on net spending in respectively research and development, with a larger effect for research grants than for development grants, but also cross-scheme effects which the authors explain by complementarity between research and development. Their results seem to indicate that subsidies, whether targeted to research or development, result in higher input additionality, especially in research activities (with substantial market failures).

4.2. Composition of R&D personnel

Graph 17 shows the evolution in the number of R&D employees in the Business Sector in Belgium, grouped by educational degree, over the period 1995-2011. The largest group of R&D personnel are employees with a university degree (excluding PhDs) or a second stage tertiary degree. The trend in the number of R&D employees with a university degree, to large extent, coincides with the evolution in R&D expenditures, increasing steadily until 2011 and then dropping dramatically until 2004. From then onwards the number started to increase gradually, reaching the 2001 level in 2010. The transitory drop in 2007 and 2008 is somewhat in contrast with the evolution in R&D expenditures. The group of R&D employees with a first stage tertiary degree also increased considerably and surpassed the group of R&D employees with at most a secondary degree, which decreased almost continuously after 2001, except a short-lived rebound in 2006. R&D employees with a PhD, although still the smallest group in 2011, is the only group to have seen its number rise even through the period 2001-2005. Starting from a relatively low level, the number of R&D employees with a PhD doubled over the period 1995-2011.

40 Ernst et al. (2014) find indications that a low tax rate on patent income appears to attract innovative projects with a high earnings and innovation potential.

Graph 17 Evolution of R&D personnel by level of education (Business Sector in Belgium)
1995-2011



Source: Belgian Office of Science Policy

Two of the schemes of partial exemption for researchers, target specific educational degrees, PhDs and civil engineers or masters. R&D active companies can autonomously decide how to spend the money they save through the partial exemption of the advance payment of the withholding tax. They have no obligation to spend it on R&D activities or on a specific group of researchers. However, the partial exemption lowers the wage costs of specific groups of workers, increasing their attractiveness relative to employees that are not eligible for support, which invokes the following question:

Do companies change the composition of their R&D personnel, in terms of educational degree, as a result of the subsidies or tax benefits they receive?

As with the variable on the orientation of R&D activities, the breakdown of R&D personnel by educational degree is only given for one of the two years covered by the biennial R&D survey, so a pseudo panel is constructed. As the estimation considers the shares of different groups of employees in total R&D personnel, Seemingly Unrelated Regression is performed

Table 32 shows the results for three categories: R&D employees with a PhD, R&D employees with a university degree or a second stage tertiary degree and R&D employees with a first stage tertiary degree.[41] Subsidies and three of the four schemes of partial exemption have a statistically significant on the share of R&D employees with a PhD. The effect of the partial exemption for researchers with a master degree is also statistically significant (at the 10% level) but negative. The latter scheme, not too

41 A seemingly unrelated regression of share equations implies that one of the share equations needs to be dropped. In Table 32, the equation of the share of R&D employees with at most a secondary degree is dropped. In an estimation in which the share of R&D employees with a first stage tertiary degree is dropped indicates that public support does not appear to have any impact on R&D employees with at most a secondary degree. Seemingly unrelated regressions with other share combinations provide similar results.

surprisingly, has a significant and relative strong positive effect on the share in R&D personnel of researchers with a university or second stage tertiary degree. None of the other support schemes has a statistically significant impact on this share. With the exception of the tax deduction of 80% of patent income, which has a positive impact (significant at 10%), subsidies and all tax incentives have a negative impact on the share of R&D employees with a first stage tertiary degree.

Table 32 Results of panel estimation of the composition of R&D personnel (2003-2011)

Dependent variable:	PhD	University - Tertiary (2^{nd} stage)	Tertiary 1^{st} stage
Explanatory variables:			
Regional subsidy	0.003 (3.26)***	-0.001 (-0.83)	-0.002 (-1.52)
Research cooperation	0.004 (2.44)**	0.000 (0.01)	-0.003 (-1.14)
Young Innovative Company	0.008 (2.92)***	0.002 (0.52)	-0.005 (-1.33)
PhDs and civil engineers	0.006 (5.04)***	-0.000 (-0.19)	-0.004 (-2.25)**
Master	-0.002 (-1.60)*	0.008 (3.58)***	-0.005 (-2.90)***
Tax credit R&D	-0.001 (-0.31)	-0.002 (-0.64)	-0.007 (-0.23)
Tax deduction 80% patent income	-0.002 (-1.12)	-0.002 (-0.57)	0.006 (1.93)**
Control variables:			
Cash flow	0.011 (2.61)***	0.004 (0.59)	-0.013 (-2.22)**
Number of employees	-0.017 (-3.21)***	-0.010 (-1.10)	0.03 (3.97)***
Age	-0.001 (-2.86)***	0.001 (1.60)*	0.000 (0.07)
Capital intensity	-0.000 (-0.02)	-0.022 (-3.52)***	0.001 (0.30)
Region dummies		Yes	
Industry (two-digit NACE)		Yes	
Year dummies		Yes	
Industry x year dummies		No	
R-squared	0.13	0.07	0.06
Number of observations		2,113	

Note: The table shows the results of seemingly unrelated regression, using the shares of specific groups of R&D personnel, grouped by educational degree.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

For the partial exemption for researchers with a PhD and civil engineer degree or researchers with a master degree, the coefficient is statistically significant. For these two schemes there are statistically significant indications that the partial exemption results in substitution of one group of R&D employees for another, in line with the specific group of researchers that the exemption targets. The substitution appears to be substantial for the partial exemption for researchers with a master degree to the extent that it may explain the finding in Table 16 that this scheme has a significant positive impact on total R&D expenditures but not on the total number of researchers, as a shift away from R&D employees with a first stage tertiary degree towards researchers with a university or second stage tertiary degree may raise R&D expenditures (wage costs) but not necessarily the total number of researchers. A regression of the average wage of researchers on the public support variables, the shares of groups of researchers and on control variables does not provide any statistically significant coefficients and therefore does not permit to corroborate the results of some scholars that part of the positive impact of public support on R&D expenditures of firms may be due to an increase in the wages of researchers (for example, Goolsbee 1998; Marey and Borghans 2000; Jaumotte and Pain 2005; Hægeland and Møen 2007b; Lokshin and Mohnen 2013).[42]

4.3. Output of R&D activities

A number of recent studies assess the potential effects of public support for R&D on output variables, such as innovation (e.g. Cappelen et al. 2008; Garcia and Mohnen 2010; patents (e.g. Czarnitzki and Lopes-Bento 2014); profits (e.g. Cappelen et al. 2008; Takalo et al. 2013) and R&D collaboration (e.g. Hottenrott and Lopes-Bento 2014).

Public support for private R&D activities is provided because it is believed that these activities have a positive impact on economic growth. Although the rationale of public support is based on the assumption that due to spillovers firms will invest less in R&D than beneficial for overall welfare, if public support raises R&D activities the question pops up whether the support eventually also affects the output of the individual companies that receive the support:

Do subsidies or tax incentives have a positive impact on the productivity of the companies that receive the support?

Given available data, the output measure considered in this section is labour productivity, measured by respectively value added per employee and cash flow per employee. To assess the potential impact of public support on productivity, R&D expenditures are considered for different groups. The first group considers R&D expenditures of firms that did not receive support in a given year. For the subsidies and tax incentives, R&D expenditures of firms that only receive that specific support are considered and the final group considers the R&D expenditures of firms that combine at least two support schemes. As R&D expenditures will not immediately result in any output, a lag of two years is considered for R&D to have any effect on productivity. The lag could of course be more substantial but given the relatively short period of public support considering much longer lags is not feasible.

The result of a regression of productivity on the support variables are reported in Table 33. To account for the known persistence in innovation in general (e.g. Peters 2009) and productivity more specifically (e.g. Raymond et al. 2010) a lag of productivity is included in the regression. The coefficients denote the elasticity of productivity with respect to the R&D activities of the specific groups of firms.[43] The coefficients are close to the 3% return of R&D investment reported by Lehto (2007) and Peters (2009). Lehto (2007) found that fixed effects tend to underestimate the impact of R&D on productivity. The results reported in Table 33 are those of a random effects specification. The coefficients of the return to R&D do not differ too much between the different groups. All coefficients are statistically significant but random effects may overestimate statistical significance.

[42] Dumont, Spithoven and Teirlinck (2014) show that accounting for changes in the composition of R&D personnel reduces the estimates of the impact of public support on the wages of researchers.

[43] To account for a possible selection bias (e.g. "picking the winner"), three inverse Mills ratios, derived from a first-step Multinomial Logit selection model are also included in the specification but not reported. None of the coefficients of these ratios is statistically significant, indicating that there appears to be no selection bias in terms of the effects of support on productivity.

Table 33 Results of a panel estimation of the effects of support for R&D on productivity (2003-2011)

	Value added per employee		Cash flow per employee
Explanatory variables:			
Productivity (t-1)	0.32 (0.00)***		0.50 (15.74)***
Firms without support (t-2)	0.03 (5.15)***		0.06 (4.35)***
Regional subsidy (t-2)	0.04 (6.02)***		0.06 (4.44)***
Research cooperation (t-2)	0.03 (4.17)***		0.07 (3.34)***
Young Innovative Company (t-2)	0.04 (4.09)***		0.07 (2.65)***
PhDs and civil engineers (t-2)	0.02 (3.77)***		0.06 (3.61)***
Master (t-2)	0.02 (3.81)***		0.06 (3.00)***
Tax credit R&D (t-2)	0.03 (4.24)***		0.06 (3.94)***
Tax deduction 80% patent income (t-2)	0.03 (4.27)***		0.06 (3.51)***
Firms that combine support schemes (t-2)	0.03 (4.25)***		0.06 (3.98)***
Control variables:			
Number of employees	-0.56 (-21.15)***		-0.08 (-4.38)***
Age	0.00 (2.52)**		0.00 (2.34)**
Capital intensity	0.03 (4.61)***		0.18 (8.17)***
Region dummies		Yes	
Industry (two-digit NACE)		Yes	
Year dummies		Yes	
Industry x year dummies		No	
R-squared	0.96		0.59
Number of observations	2,890		2,696

Note: The table shows the results of a random effects regression of labour productivity (two alternative measures) on public support and control variables, including three invers Mills ratios derived from a first step selection model as reported in Table 21.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

For the specification with value added per employee as the productivity measure, significance tests of the difference between coefficients of different groups indicate that the group of firms that only receive direct support (regional subsidies) have a statistically significant higher return on R&D than firms that do not receive any support. Firms that only receive partial exemption for researchers with a PhD or civil engineer degree and for researchers with a master degree apparently have a statistically significant lower return on R&D than firms that receive no support. The same applies to firms that combine at least two of the support schemes although this difference is only significant at the 10% level.

In a fixed effects specification, only the return on R&D for regional subsidies and the partial exemption for Young Innovative Companies is positive and statistically significant. As the specification contains a lagged dependent variable, the panel specification may be biased. A two-step GMM estimation (not reported) provides similar results as the random effects specification but as in section 3.9, the specification performs poorly on most tests (e.g. rejection instrument validity). Following Lehto (2007), spillover variables were constructed, reflecting R&D expenditures in the same industry as the firm (intra-sector) and R&D expenditures in other industries (inter-sector). The two spillover variables are split into two groups, firms that received support and firms that received no support. The results of this specification (not reported) suggest that, for the specification with value added per employee as productivity measure, inter-sector spillovers are positive and statistically significant but appear to be higher for spillovers from firms that did not receive any support than for spillovers from firms that received support.[44] For intra-sector spillovers only the coefficient of firms that received support (two-year lag) is statistically significant but with a surprising negative sign.

For the specification with cash flow per employee as productivity measure, again all coefficients are highly significant. The average rate of return appears to be higher than for value added per employee. The differences across groups are not very large and actually no difference between the group of firms that do not receive any support and the groups of firms that receive support is statistically significant. A two-step system GMM specification to account for the possible bias due to the inclusion of a lagged dependent variable (not reported) provides similar results and even has reasonable specification results (e.g. no rejection of the null hypothesis of the Hansen over-identifying restrictions test). A specification with spillover variables (not reported) only provides a statistically significant coefficient for intra-sector spillovers from firms that received support (two-year lag) which, as in the value added per employee specification, has a negative sign.

A more detailed elaboration of spillovers is beyond the scope of this paper but – given their importance for the rationale of public support of private R&D – seems recommendable for future evaluations.

44 The spillovers are also considered with a two-year lag.

5. Conclusions

Considering the results of the different estimation procedures presented in this paper, a statistically significant positive impact of the partial exemption from advance payment of the withholding tax on the wages of researchers with a master degree and direct support (regional subsidies) appears to be the most robust finding. The effect of the partial exemption for researchers with a master degree is more important in terms of R&D expenditures than in terms of the total number of R&D employees as this support scheme appears to result in a substitution of researchers with a university or second-stage tertiary degree for researchers with a PhD and especially for R&D personnel with a first-stage tertiary degree. The partial exemption from advance payment of the withholding tax on the wages of researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineering degree also appears to result in substitution, of researchers with a PhD for R&D personnel with a first stage tertiary degree, but the substitution is only partial as this support scheme has a statistically significant impact on the total number of employees involved in R&D.

The statistically significant positive impact of the partial exemption from advance payment of the withholding tax on the wages of researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineer degree and on the wages of R&D personnel of companies that cooperate in research with a university, higher education institution or scientific institution in the baseline specification is not confirmed by alternative estimates that account for selection and endogeneity.

There are few, if any, indications that the tax benefit (partial exemption) for Young Innovative Companies, the tax credit for R&D investment and the tax deduction of 80% of gross patent income have a statistically significant effect on private R&D expenditures. On the other hand, the partial exemption for Young Innovative Companies, the tax deduction of 80% of gross patent income and regional subsidies seem to induce firms to shift their R&D activities away from development towards research (basic or applied). This result seems relevant in view of the recent reports – to some extent corroborated in this paper – that large companies increasingly shy away from scientific research which they leave up to small start-ups and the rationale of public support which especially holds for risky projects such as basic and applied research.

Results reported in this paper confirm the indications in the first evaluation that the combination of different support schemes seems to reduce the effectiveness of public support. This appears to be the case for firms that combine direct support (subsidies) by the regions with a partial exemption from advance payment of the withholding tax for researchers with a master degree but also for several combinations of federal tax incentives. The reduction in effectiveness is substantial for firms that combine more than two support schemes.

In terms of potential effects of public support for R&D on output (value added per employee), results suggest that firms that only receive direct support (regional subsidies), have a higher return from R&D than firms that do not receive any support. For the other schemes no statistically significant effects are found.

There is a lot of heterogeneity across firms and industries in the impact of public support. For most support schemes, the effectiveness of support does not seem to be constant but appears to decrease, though not always monotonically, as the rate of subsidization increases.

Although simple rules are recommendable for any government policy, introducing some degree of heterogeneity – obviously in line with EU rules of state aid – seems warranted to increase the effectiveness of the tax incentives and to constrain the rising budgetary costs.

Given the relatively large number of different existing tax incentives for R&D, some streamlining could be considered. Given that the rate of exemption is the same for all schemes of partial exemption from advance payment of the withholding tax, a separate scheme for researchers with a PhD in (applied) science or (veterinary) medicine or with a civil engineer degree and for researchers with a master degree seems dispensable as the first list of degrees is – by definition – implied by the latter.

Although no indications are found that public support for R&D raised the average wages of researchers in the period 2003-2011, it is important to assure that a rising demand for R&D personnel is met with a sufficient supply of qualified researchers.

References

Achen, C. 2001. Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables, CSDP and University of Michigan.

Aerts, K., Czarnitzki, D. 2004. Using innovation survey data to evaluate R&D policy: The case of Belgium. ZEW - Centre for European Economic Research Discussion Paper No. 04-055.

Aghion, P., Askenazy, P., Berman, N., Cette, G., Eymard, L., 2012. Credit Constraints and the Cyclicality of R&D Investment: Evidence from France, Journal of the European Economic Association 10(5), 1001-1024.

Akcigit, U., Kerr, W.R. 2010. Growth through Heterogeneous Innovations, NBER Working Paper 16443.

Akcigit, U., Hanley, D., Serrano-Velarde, N. 2014. Back to Basics: Basic Research Spillovers, Innovation Policy and Growth, NBER Working Paper 19473.

Almus, M., Czarnitzki, D. 2003. The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany, Journal of Business & Economic Statistics 21(2), 226-236.

Anderson, T., Hsiao, C. 1981. Estimation of Dynamic Models with Error components, Journal of the American Statistical Association 76(375), 598–606.

Angrist, J.D., Pischke, J.-S. 2009. Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.

Angrist, J.D., Pischke, J.-S. 2015. Mastering ’Metrics: The Path from Cause to Effect, Princeton University Press.

Antonelli, C., Crespi, F. 2013. The "Matthew effect" in R&D public subsidies: The Italian evidence, Technological Forecasting and Social Change 80(8), 1523–1534.

Arora, A., Belenzon, S, Patacconi, A. 2015. Killing the Golden Goose? The Decline of Science in Corporate R&D, NBER Working Paper No. 20902.

Arellano, M., Bond, S. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58, 277–297.

Arellano, M., Bover, O. 1995. Another look at instrumental variables estimation of error-component models, Journal of Econometrics 68, 29–51.

Arqué-Castells, P. 2013. Persistence in R&D performance and its implications for the granting of subsidies, Review of Industrial Organization 43(3), 193-220.

Arqué-Castells, P., Mohnen, P. 2015. Sunk costs, extensive R&D subsidies and permanent inducement effects, Review of Industrial Organization, in press.

Baltagi, B.H., 2005. Econometric Analysis of Panel Data, 3rd. Edition, John Wiley & Sons, Chichester.

Baltagi, B.H., Griffin, J.M., 1997. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline, Journal of Econometrics 77(2), 303–327.

Barlevy, G. 2007. On the Cyclicality of Research and Development, American Economics Review 97(4), 1131-1164.

Baum, C.F. (2013), Dynamic Panel Data estimators, Boston College, slides.

Becker, B. 2013. The Determinants of R&D Investment: A Survey of the Empirical Research, Discussion Paper Series 2013_09, Department of Economics, Loughborough University, revised Sep 2013.

Bia, M., Mattei, A. 2008. A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score, The Stata Journal 8(3), 354-373.

Blanes, J.V., Busom, I. 2004. Who participates in R&D subsidy programs?: The case of Spanish manufacturing firms, Research Policy 33, 1459–1476.

Blundell, R., Bond, S. 1998. Initial conditions and moment restrictions in dynamic panel-data models, Journal of Econometrics 87, 115–143.

Blundell, R.W., Costa Dias, M. 2009. Alternative Approaches to Evaluation in Empirical Microeconomics, Journal of Human Resources 44(3), 565-640.

Bodas Freitas, I., Castellacci, F., Fontana, R., Malerba, F., Vezzulli, A. 2015. The additionality effects of R&D tax credits across sectors: A cross-country microeconometric analysis, Working Papers on Innovation Studies 20150424, Centre for Technology, Innovation and Culture, University of Oslo.

Bogliacino, F. Pianta, M. 2015. The Pavitt Taxonomy, Revisited. Patterns of Innovation in Manufacturing and Services, Documentos de Trabajo – Escuela de Economía 012631.

Bound, J., Jaeger, D.A., Baker, R.M. 1995. Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak, Journal of the American Statistical Association 90(430), 443-450.

Brown, J.R., Fazzari, S.M., Petersen, B.C. 2009. Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom, Journal of Finance 64(1), 151-185.

Brown, J.R., Martinsson, G., Petersen, B.C. 2012. Do financing constraints matter for R&D?, European Economic Review 56(8), 1512-1529.

Busom, I. 2000. An Empirical Evaluation of the Effects of R&D Subsidies, Economics of Innovation and New Technology 9(2), 111-148.

Busom, I., Corchuelo, B., Martínez-Ros, E. 2011. Tax incentives and direct support for R&D: what do firms use and why?, Business Economics Working Papers id-11-03, Universidad Carlos III, Instituto sobre Desarrollo Empresarial (INDEM).

Busom, I., Corchuelo, B., Martínez-Ros, E. 2015. Dynamics of firm participation in R&D tax credit and subsidy programs, Working Paper 15.03, Department of Applied Economics at Universitat Autonoma of Barcelona.

Busom, I., Fernández-Ribas, A. 2007, Do R&D programs of different Government levels overlap in the European Union?, Instituto de Estudios Fiscales.

Caliendo, M., Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching, Journal of Economic Surveys 22 (1), 31-72.

Cameron, A.C., Trivedi, P.K. 2010. Microeconometrics Using Stata (Revised Edition), Stata Press.

Cantner, U., Kösters, S. 2012. Picking the winner? Empirical evidence on the targeting of R&D subsidies to start-ups, Small Business Economics 39(4), 921-936.

Cappelen,Å., Raknerud, R., Rybalka, M. 2008. The effects of R&D tax credits on patenting and innovations, Discussion Paper 565, Research Department of Statistics Norway.

Castellacci, F., Lie, C.M. 2015. Do the effects of R&D tax credits vary across industries? A meta-regression analysis, Research Policy 44(4), 819-832.

Cerulli, G. 2010. Modelling and Measuring the Effect of Public Subsidies on Business R&D: A Critical Review of the Econometric Literature, The Economic Record 86 (274), 421-449.

Cerulli, G. 2014. CTREATREG: Stata module for estimating dose-response models under exogenous and endogenous treatment, Working paper series, Ceris-CNR 05-2014.

Cerulli, G., Potì, B. 2010, The differential impact of privately and publicly funded R&D on R&D investment and innovation: The Italian case, Working Papers10, Doctoral School of Economics, Sapienza University of Rome.

Cerulli, G., Potì, B. 2014. The Impact of Public Support Intensity on Business R&D: Evidence from a Dose-Response Approach, ERSA conference papers ersa14p625, European Regional Science Association.

Chang A.C. 2012. Tax Policy Endogeneity: Evidence from R&D Tax Credits. Job Market paper, University of California – Irvine.

Cheng, T.C., Trivedi, P.K. 2014. Attrition Bias in Panel Data: A Sheep in Wolf's Clothing? A Case Study Based on the MABEL Survey, Melbourne Institute Working Paper No. 14/14, Melbourne Institute of Applied Economics and Social Research.

Cincera, M., Ravet, J., Veugelers, R. 2014. R&D financing constraints of young and old innovation leaders in the EU and the US, WP 2014 – 008, The International Centre for Innovation, Technology and Education Studies (iCite), Solvay Brussels School of Economics and Management.

Clausen, T. 2008. Do subsidies have positive impacts on R&D and innovation activities at the firm level?, Working Paper on Innovation Studies 20070615, Centre for Technology, Innovation and Culture, University of Oslo.

Coad, A., Rao, R. 2010. Firm growth and R&D expenditure, Economics of Innovation and New Technology 19(2), 127-145.

Cohen, W.M., Klepper, S. 1996. A Reprise of Size and R&D, Economic Journal 106(437), 925-951.

Corchuelo, B. and E. Martínez-Ros (2009), The Effects of Fiscal Incentives for R&D in Spain, Business Economics Working Papers wb092302, Universidad Carlos III, Departamento de Economía de la Empresa.

Crown, W.H., Henk H.J., Vanness D.J. 2011. Some cautions on the use of instrumental variables estimators in outcomes research: how bias in instrumental variables estimators is affected by instrument strength, instrument contamination, and sample size, Value Health 14(8), 1078-1084.

Czarnitzki, D., Lopes-Bento, C. 2013. Value for money? New microeconometric evidence on public R&D grants in Flanders, Research Policy 42(1), 76-89.

Czarnitzki, D., Lopes-Bento, C. 2014. Innovation Subsidies: Does the Funding Source Matter for Innovation Intensity and Performance? Empirical Evidence from Germany, Industry and Innovation 21(5), 380-409.

da Silva, A.M.N.M. 2014. The Determinants of Participation in R&D Subsidy Programmes: Evidence from Firms and S&T Organisations in Portugal, Master Dissertation in Innovation Economics and Management, Faculdade de Economia da Universidade do Porto.

De Blasio, G., Fantino, D., Pellegrini, G. 2014. Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds, Industrial and Corporate Change, doi: 10.1093/icc/dtu027, first published online: October 6, 2014.

Dickey, D. A., Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427–431.

Driscoll, J. C., Kraay, A.C. 1998. Consistent covariance matrix estimation with spatially dependent panel data, Review of Economics and Statistics 80, 549-560.

Drukker, D.M. 2008. Econometric analysis of dynamic panel-data models using Stata, Summer North American Stata Users Group meeting.

Dubin J.A., McFadden, D.L. 1984. An Econometric Analysis of Residential Electric Appliance Holdings and Consumption, Econometrica 52(2), 345-362.

Dumont, M. 2013. The impact of subsidies and fiscal incentives on corporate R&D expenditures in Belgium (2001-2009), Working Paper 01-13, Federal Planning Bureau.

Dumont, M., Spithoven, A., Teirlinck, P. Public support for R&D and the educational mix of R&D employees, Working Paper 14-08, Federal Planning Bureau, Belgium.

Einiö, E. 2014. R&D Subsidies and Company Performance: Evidence from Geographic Variation in Government Funding Based on the ERDF Population- Density Rule, Review of Economics and Statistics 96(4), 710–728.

Ernst, C., Richter, K., Riedel, N. 2014. Corporate taxation and the quality of research and development, International Tax and Public Finance 21(4), 694-719.

Evers, L., Miller, H., Spengel, C. 2015. Intellectual property box regimes: effective tax rates and tax policy considerations, International Tax and Public Finance 22(3), 502–530.

Fabrizio, K.R., Tsolmon, U. 2014. An Empirical Examination of the Procyclicality of R&D Investment and Innovation, Review of Economics and Statistics 96(4), 662-675.

Garcia, A., Mohnen, P. 2010. Impact of government support on R&D and innovation, UNU-MERIT Working Paper Series 034, Maastricht.

Globerman, S. 1973. Market Structure and R&D in Canadian Manufacturing Industries, Quarterly Review of Economics and Business 13, 59-67.

Geroski, P. A., Van Reenen, J., Walters, C. F. 1997. How persistently do firms innovate?, Research Policy 26(1), 33-48.

Gonzalez, X., Jaumandreu, J., Pazo, C. 2005. Barriers to Innovation and Subsidy Effectiveness, RAND Journal of Economics 36(4), 930-949.

Goolsbee A. 1998. Does Government R&D Policy mainly benefit Scientists and Engineers?, American Economic Review 88(2), 292-302.

Görg, H., Strobl, E. 2007. The Effect of R&D Subsidies on Private R&D, Economica 74(294), 215-234.

Greene, W.H. 2000, Econometric Analysis (fourth edition), Prentice-Hall, New Jersey.

Griffith, R., Miller, H., O'Connell, M. 2014. Ownership of intellectual property and corporate taxation, Journal of Public Economics 112, 12–23.

Griliches, Z., Hausman, J.A. 1986. Errors in variables in panel data, Journal of Econometrics 31(1), 93-118.

Hægeland, T., Møen, J. 2007 a. Additionality of R&D subsidies: A comparison between Tax Credits and Direct Grants, Statistics Norway Reports 2007/45.

Hægeland T, Møen J. 2007 b. Input Additionality in the Norwegian R&D Tax Credit Scheme. Statistics Norway, Oslo.

Hall, B.H., Lerner, J. 2010. The Financing of R&D and Innovation, in: B. H. Hall and N. Rosenberg (eds.), Handbook of the Economics of Innovation, 609-639.

Haskel, J., Goodridge, P., Corrado, C. 2011. Constructing a price deflator for R&D: calculating the price of knowledge investments as a residual, Working Papers 9028, Imperial College, London.

Heckman, J.J. 1979. Sample selection Bias as a Specification Error, Econometrica 47(1), 153-161.

Heckman, J., Ichimura, H., Smith, J., Todd, P. 1998. Characterizing selection bias using experimental data, Econometrica 66(5), 1017-1098.

Henningsen, M., Hægeland, T., Møen, J. 2015. Estimating the Additionality of R&D Subsidies Using Proposal Evaluation Data to Control for Research Intentions, The Journal of Technology Transfer 40(2), 227-251.

Hoechle, D. 2007. Robust standard errors for panel regressions with cross–sectional dependence, The Stata Journal 7(3), 281-312.

Holtz-Eakin, D., Newey, W., Rosen, H.S. 1988. Estimating Vector Autoregressions with Panel data, Econometrica 56(6), 1371–1395.

Hottenrott, H., Lopes-Bento, C., 2014. (International) R&D collaboration and SMEs: The effectiveness of targeted public R&D support schemes, Research Policy 43(6), 1055-1066.

Hottenrott, H., Lopes-Bento, C., Veugelers, R. 2014, Direct and Cross-Scheme Effects in a Research and Development Subsidy Program, ZEW Discussion Paper 14-10.

Huergo, E., Moreno, L. 2011. Does history matter for the relationship between R&D, innovation, and productivity?, Industrial and Corporate Change 20(5), 1335-1368.

Huergo, E., Moreno, L. 2014. National or international public funding? Subsidies or loans? Evaluating the innovation impact of R&D support programmes, MPRA Paper 54218, University Library of Munich, Germany.

Hussinger, K. 2008. R&D and subsidies at the firm level: an application of parametric and semiparametric two-step selection models, Journal of Applied Econometrics 23(6), 729-747.

Imbens, G.W. 2014. Instrumental Variables: An Econometrician's Perspective, Statistical Science 29 (3), 323-358.

Jaffe, A. 2002. Building Programme Evaluation into the Design of Public Research-Support Programmes, Oxford Review of Economic Policy 18(1), 22-34.

Jaumotte F, Pain N. 2005. From Ideas to Development: The Determinants of R&D and Patenting. OECD Economics Department Working Paper No. 457, OECD Publishing.

Kaiser, B. 2014. RHAUSMAN: Stata module to perform a (cluster-) robust Hausman test, University of Bern.

Kauko, K., 1996. Effectiveness of R&D subsidies - a sceptical note on the empirical literature, Research Policy 25(3), 321-323.

Keele, L.J., Kelly, N.J. 2006. Dynamic Models for Dynamic Theories: The Ins and Outs of LDVs, Political Analysis 14(2), 186-205.

Kerr, W.R., Nanda, R. 2014. Financing Innovation, Harvard Business School Working Paper, No. 15-034.

Klette, T.J., Kortum, S. 2004. Innovating Firms and Aggregate Innovation, Journal of Political Economy 112(5), 986-1018.

Klette, T.J., Møen, J., 2012. R&D investment responses to R&D subsidies: A theoretical analysis and a microeconometric study, World Review of Science Technology and Sustainable Development 09(2/3/4), 169 – 203.

Klette, T.J., Møen, J., Griliches, Z. 2000. Do subsidies to commercial R & D reduce market failures? Microeconometric evaluation studies, Research Policy 29, 471–495.

Lehto E. 2007. Regional Impact of Research and Development on Productivity, Regional Studies 41(5), 623-638.

Lichtenberg, F.R. 1984. The Relationship between Federal Contract R&D and Company R&D, American Economic Review 74(2), 73-78.

Lichtenberg, F.R. 1988. The Private R&D Investment Response to Federal Design and Technical Competitions, American Economic Review 78(3), 550-559.

Lokshin, B., Mohnen, P. 2007. Measuring the Effectiveness of R&D tax credits in the Netherlands, UNU-MERIT Working Paper 2007-025, Maastricht.

Lokshin B, Mohnen P. 2013. Do R&D tax incentives lead to higher wages for R&D workers?: Evidence from the Netherlands, Research Policy 42, 823–830.

Lööf, .H, Heshmati, A. 2005. The Impact of Public Funds on Private R&D Investment: New Evidence from a Firm Level Innovation Study, Discussion Paper 11862, MTT Agrifood Research Finland.

Máñez, J.A., Rochina-Barrachina, M.E, Sanchis, A., Sanchis, J.A. 2009. The Role of Sunk costs in the Decision to invest in R&D, Journal of Industrial Economics 57(4), 712-735.

Marey, P., Borghans L. 2000. Wage Elasticities of the Supply of R&D Workers in the Netherlands. ROA report R-2000/6E, Maastricht.

Marino, M., Parrotta, P., Sala, D. 2010. New Perspectives on the Evaluation of Public R&D Funding, Working Papers 11-2, University of Aarhus, Aarhus School of Business, Department of Economics.

Meeusen W., Janssens W. 2001. Substitution versus additionality: econometric evaluation by means of micro-economic data of the efficacy and efficiency of R&D subsidies to firms in the Flemish Region, CESIT paper 2001/01.

Mulkay, B., Mairesse, J. 2013. The R&D tax credit in France: assessment and ex ante evaluation of the 2008 reform, Oxford Economic Papers 65(3), 746-766.

Nelson, C.R., Startz, R. 1988. The Distribution of the Instrumental Variables Estimator and its t-ratio when the Instrument is a Poor One, NBER Technical Working Papers 0069.

Nelson, C.R., Startz, R. 1990. Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator, Econometrica 58(4), 967-976.

Nichols, A. 2007. Causal inference with observational data, The Stata Journal 7(4), 507-541.

Nickell, S.J. 1981. Biases in Dynamic Models with Fixed Effects, Econometrica 49(6), 1417-1426.

OECD 2013. Science, Technology and Industry Scoreboard.

OECD 2015. OECD/G20 Base Erosion and Profit Shifting Project, Action 5: Agreement on Modified Nexus Approach for IP Regimes.

Ouyang, M. 2011. On the Cyclicality of R&D, Review of Economics and Statistics 93(2), 542-553.

Park, S. 2010. R&D Intensity and Firm Size Revisited, Job Market Paper, University of California Los Angeles.

Pavitt, K. (1984). Patterns of technical change: towards a taxonomy and a theory, Research Policy 13, 343-374.

Peters, B. 2009. Persistence of innovation: stylised facts and panel data evidence, The Journal of Technology Transfer 34(2), 226-243.

Rao N. 2013. Do Tax Credits Stimulate R&D Spending? The Effect of the R&D Tax Credit in its First Decade. NYU's Wagner School of Public Service.

Raymond, W., Mohnen, P., Palm, F., Schim van der Loeff, S., 2010. Persistence of innovation in Dutch manufacturing: Is it spurious?, Review of Economics and Statistics 92, 495–504.

Robbins, R., Belay, O., Donahoe, M., Lee, J. 2012. Industry-level Output Price Indexes for R&D: An Input-cost Approach with R&D Productivity Adjustment, BEA Working Paper.

Roodman, D. 2007. A Note on the Theme of Too Many Instruments, Working Paper 125, Center for Global Development.

Roodman, D. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata, The Stata Journal 9(1), 86-136.

Rosenbaum, P.R., Rubin, D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika 70(1), 41-55.

Social Research and Demonstration Corporation 2007. The Use of Randomized Field Experiments to Assess the Effectiveness of Government Programs to Support Business Innovation: A Feasibility Study, report submitted to Industry Canada.

Stancík, J., Biagi, F. 2012. Characterizing the Evolution of the EU-US R&D Intensity Gap using Data from Top R&D Performers, Ecomod paper 4207.

Streicher, G., Schibany, A., Gretzmacher, N. 2004. Input Additionality Effects of R&D Subsidies in Austria: Empirical evidence from firm-level panel data, Institute of Technology and Regional Policy-Joanneum Research, Vienna.

Suetens S. 2002. R&D subsidies and production effects of R&D personnel: evidence from the Flemish Region, CESIT discussion paper 2002/03.

Symeonidis, G. 1996. Innovation, Firm Size and Market Structure: Schumpeterian Hypotheses and Some New Themes, OECD Economics Department Working Paper 161.

Takalo, T., Tanayama, T., Toivanen, O. 2013. Estimating the Benefits of Targeted R&D Subsidies, Review of Economics and Statistics 95(1), 255-272.

Thakor R. T., Lo, A.W. 2015. Competition and R&D Financing Decisions: Theory and Evidence from the Biopharmaceutical Industry, NBER Working Paper 20903.

Triguero, A., Córcoles D., Cuerva, M. 2014. Persistence of innovation and firm's growth: evidence from a panel of SME and large Spanish manufacturing firms, Small Business Economics 43(4), 787-804.

Varian, H.R. 2015. Causal Inference in Social Science: An elementary introduction, mimeo.

Wallsten, S. 2000. The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program, RAND Journal of Economics 31(1), 82-100.

Wilkins, A.S. 2014. To Lag or Not to Lag? Re-evaluating the Use of Lagged Dependent Variables in Regression Analysis, Stanford University, Working Paper.

Windmeijer, F. 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of Econometrics 126, 25-51.

Woerter, M. 2014. Competition and Persistence of R&D, Economics of Innovation and New Technology 23(5-6), 469-489.

Wolff, G.B., Reinthaler, V. 2008. The effectiveness of subsidies revisited: Accounting for wage and employment effects in business R&D, Research Policy 37(8), 1403-1412.

Zivot, E., Jiahui J. 2002. Modeling Financial Time Series with S-PLUS, Springer-Verlag.

Zúñiga-Vicente, J.A., Alonso-Borrego, C., Forcadell, F.J., Galán, J.I. 2014. Assessing the Effect of Public Subsidies on Firm R&D Investment: A Survey, Journal of Economic Surveys 28(1), 36-67.

Annex

Annex 1: Technical details on estimation procedures

In this annex the different estimation procedures that are used in this paper to account for the selection and endogeneity of public support for R&D are discussed in more technical detail than in the main text. The elaboration by Angrist and Pishke (2009, 2015) of the difficulties in establishing a causal effect is taken as the lead in this discussion.

The causal effect

Angrist and Pishke (2009, 2015) point out that the most credible assessment of the causal effect of a given treatment results from random assignment. In the case of public support for R&D activities of private companies, this would imply that an agency grants subsidies or tax benefits fully randomly, in effect, irrespective of any firm characteristic or the quality of a project proposal. Such an experiment would permit to evaluate the evolution in the R&D activities of the group of firms that receives support (treatment group) to the evolution in the R&D activities of the group of firms that do not receive support (control group). Adapting the example provided by Angrist and Pishke (2009: pp. 9-11) to the context of public support for R&D, the R&D expenditures of firm *i* if it receives support (D_i=1) are denoted as Y_{1i} and if it does not receive support (D_i=0) as Y_{0i}. The observed outcome Y_i can be written as:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i \qquad \text{A.1}$$

The causal effect of public support is given by the difference between the R&D expenditures of a firm if it receives support and its R&D expenditures if it does not receive support, Y_{1i} -Y_{0i}. As a firm can at a given moment in time either receive support or not, only one option can be observed and the other (counterfactual) cannot. However, the R&D expenditures of a group of firms that receive support (D_i=1) can be compared to the outcome of a group of firms that do not receive support (D_i=0):

$$\overbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}^{Observed\ difference\ in\ R\&D\ exependitures} = \overbrace{(E[Y_{1i} - Y_{0i}|D_i = 1])}^{Average\ treatment\ effect\ on\ firms\ with\ support}$$
$$+ \overbrace{(E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0])}^{Selection\ bias} \qquad \text{A.2}$$

The first term in round brackets on the right-hand side is the causal effect of public support that we are interested in, the average treatment effect on the treated. As the counterfactual for an individual firm does not exist, this term cannot observed. The second term on the right hand side, provides an indication of the selection bias and can also – lack of the counterfactual – not be observed. However, if firms are randomly assigned, D_i is independent of potential R&D expenditures so that $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ and the last term on the right-hand side, the selection bias, disappears. The observed difference in the R&D expenditures between firms with support and those without support equals the causal effect of support. Random assignment is what Angrist and Pishke (2009) consider to be the "experimental ideal". Unfortunately – from the perspective of the evaluator – it is very rare that public support is granted to companies in a fully randomized way.

Jaffe (2002) pleads for the use of randomization in granting subsidies. This could be operationalized by dividing applicants, based on the quality of proposals, into groups, for example high priority; marginal and rejections. The probability to receive a subsidy could be higher for the high priority proposals than for the marginal proposals but randomization implies that some high priority proposals need to be refused funding. Jaffe, aware of the political reticence with regard to randomized funding, considers an alternative that consists in granting all high priority applicants support and to provide support to the marginal applicants in a randomized way (see, for example, the 2007 report by the Social Research and Demonstration Corporation on randomized experiments to assess the effectiveness of public support for innovation). If randomized experiments to assess the impact of subsidies are rare, it seems hardly feasible to run an experiment that would consist in assigning randomly subsidies as well as six different tax benefits as in the case in the evaluation discussed in this paper.

Absent the experimental ideal of random assignment[45], Angrist and Pishke (2009, 2015) consider four alternative methods that can be used to investigate causal effects[46]:

- Regression
- Instrumental variables
- Regression discontinuity
- Differences-in-differences

These four methods consist in accounting for the selection bias and endogeneity in a context in which random assignment is not applicable and assessment is based on observational (post-treatment) data. Causal inference of these methods relies on the *conditional independence assumption* (CIA) or the *selection-on-observables* assumption, i.e. all potential covariates that may affect (self-) selection are known and observed. If the CIA holds, the selection bias can be removed by controlling for observed covariates and a causal interpretation can be justified (Angrist and Pishke 2009: pp. 39-40). Angrist and Pishke (2009) argue that including covariates as control variable increases the plausibility of the CIA.

In the following section we discuss the first two of the four procedures listed by Angrist and Pishke (2009, 2015) that are used in this paper. Regression discontinuity is not discussed as this approach requires some threshold or discontinuity that does not seem available in the data set used in this paper (see De Blasio et al. 2014) for an example of a regression discontinuity applied in the evaluation of R&D subsidies in Italy, using an unexpected shortage of public money as discontinuity). Differences-in-differences (DiD) is mostly applied for cross-sectional data or when no data over a long period of time are available. The fixed effects regression, as discussed in the next section, can be considered as a generalization of DiD in a panel context (Heckman et al. 1998; Angrist and Pishke 2009; Klette et al. 2000). Klette and Moen (2012) argue that when data for more than two years are available, and the amount of support is known, a fixed effects specification is to be preferred to a DiD estimator.

[45] Angrist and Pishke (2009: p.15) point at the limitations of real-life randomized trials.

[46] See for similar arguments, Jaffe (2002); Nichols (2007) and Varian (2015). Cerulli (2010) also considers more recent dynamic models of imperfect competition, which are beyond the scope of this paper

Panel regression

Angrist and Pishke (2009: pp. 26-27) argue that regression estimates provide a valuable baseline for most empirical research because it is closely linked to the Conditional Expectation Function (CEF), which provides a natural summary of empirical relationships. The matched data cover firms over a period of time, which permits to construct a panel. Panel regression, with covariates denoting potential determinants of private R&D activities, is considered as the baseline specification in this paper.[47] A panel has two dimensions, the cross-section dimension *i* (firms in this paper) and the time dimension *t*:

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \qquad \text{A.3}$$

The dependent variable of firm *i* in year *t* is denoted by y_{it}. X'_{it} is the matrix of right-hand-side variables (ʻ denotes the transpose of a matrix), α is a scalar (intercept) and β is the vector (column matrix) of the coefficients (parameters) to be estimated. Specification A.3 assumes that the intercept and the slope coefficients β are common to all individuals *i*. In a panel specification, heterogeneity across individuals (firms) is mostly modelled through the residual term:

$$u_{it} = \alpha_i + v_{it} \qquad \text{A.4}$$

Unobservable time-fixed firm-specific effects are denoted by α_i and v_{it} is the residual term. Fixed effects reflect possible factors that may affect the level of the dependent variable (R&D expenditures) but do not change much over time, such as the ability of managers and researchers, niche-specific technical opportunities or ownership structure. Lichtenberg (1984) pointed out that heterogeneity across firms that is not observed could be captured by firm fixed (time-invariant) effects if the unobserved heterogeneity does not change much over time. In a fixed effects estimation α_i are assumed to be correlated with v_{it}, assumed independent and identically distributed $N(0, \sigma^2)$. Fixed effects (FE) can be estimated by OLS considering a specification that includes a dummy for each individual (Least Squares Dummy Variable, LSDV). An alternative, which provides the same results, is to perform OLS on a specification in which all variables are included after a within transformation, which consists in subtracting the average of a variable across years for each individual firm, for example a right-hand-side variable *x*:

$$x_i^{Within} = x_{it} - \overline{x_\iota} \qquad \text{A.5}$$

The advantage of the within transformation is that in a large panel, with many individuals, the OLS matrix does not become intractably large as is the case if a large number of dummies are included, leading to a substantial loss of degrees of freedom. If the traditional assumption for v_{it} holds, LSDV (within) is the best linear unbiased estimator. In a short panel (many firms, few years), as is the case in this paper, the FE estimator of β is consistent as the number of individuals N becomes very large (moves towards infinity). The estimator of the firm-fixed effects is not consistent as the number of these effects also increases as N increases. If the true model is one with time-invariant individual-specific effects, OLS will provide biased estimates.

An alternative to the fixed effects (FE) specification is random effects (RE), which assumes the firm-specific effects to be random ($\alpha_i \sim IID(0, \sigma_\alpha^2)$ and therefore, in contrast with FE, to be independent from

[47] This section is to a large extent based on Baltagi (2005).

$v_{it} \sim IID(0, \sigma_v^2)$. Random effects may be useful if individuals are drawn randomly from a population. A Hausman test can be used to decide between FE and RE. The null hypothesis of the test is that RE is fully efficient. Rejection of the null hypothesis would favour FE over RE. However, if heteroskedasticity or serial correlation is present in the panel, as is often the case, the RE estimator is not fully efficient and a traditional Hausman test may result in over-rejection of the null hypothesis. In this paper a robust Hausman test is applied (Stata procedure RHAUSMAN proposed by Kaiser 2014):

$$H= (B_{RE}-B_{FE})' * [V_{Bootstrapped}(B_{RE}-B_{FE}]^{(-1)} * (B_{RE}-B_{FE}) \sim chi2(k) \quad A.6$$

B_{RE}- is the vector of coefficient estimates from the RE estimation and B_{FE} is the vector of FE coefficient estimates. By using the cluster-robust covariance matrix $V_{Bootstrapped}(B_{RE}-B_{FE})$, the robust Hausman test does not require RE (or FE) to be fully efficient under the null hypothesis. A rejection of the robust Hausman tests also suggests to favour the results of a FE estimation over the results of a RE estimation because, if FE is the true model, RE does not provide consistent estimates.

Cerulli (2010) points out that if suitable covariates are considered, exogeneity of the right-hand-side variables is restored. As pointed out before, causal inference of regression techniques depends on the conditional independence assumption (CIA) which implies that all potential covariates that may affect (self-) selection are known and observed. A selection bias may result from heterogeneity across firms in applying for subsidies or tax benefits (self-selection) but also from specific goals of agencies that provide public support, for example because of a focus on specific technological domains industries or a government policy favouring SMEs that face more credit constraints than larger companies (e.g. Klette et al. 2000; Jaffe 2002; Blanes and Busom 2004; Huergo and Moreno 2014).

Heckman (1979) proposed a two-step procedure to account for the selection problem. In a first step, a binary variable that reflects a given choice or status (support or no support) is regressed on potential determinants. Information on the probability of an individual to have a given status (receive support or not) is subsequently used in the second step which consists in the regression of the structural specification, to account for a possible bias due to the fact that support is not applied for and granted in a random way. The fact that we consider subsidies granted by regional agencies as well as tax benefits provided by the federal government implies that we need to account for different selection modes. To receive a subsidy, companies need to apply at the funding agency. Regional subsidies are subject to selection due to factors that may incite companies to apply for a subsidy (self-selection) as well as due to the specific criteria that the agencies consider in granting subsidies. As tax benefits are, in contrast with subsidies, not granted through competition, the reasons for companies to apply for tax benefits may differ from the reasons to apply for subsidies. Agencies that have to decide whether tax benefits can be granted, have to check whether the necessary eligibility conditions are fulfilled. To account for potential differences in the granting of subsidies and tax benefits a multiple logit selection specification is considered in the first step of the two-step Heckman procedure. A number of previous studies on public support for R&D also used a multiple selection specification (for example, Huergo and Moreno 2014). Multinomial logit provides estimates of the extent to which right-hand-side variables explain the probability of a given company to receive public support. Four categories are considered: 1 (company receives no support for R&D); 2 (company receives a subsidy but no tax benefit); 3 (company receives a tax benefit but no subsidy); 4 (company receives a subsidy as well as a tax benefit). The probability to be in category i is given by:

$$\Pr(y = i) = \frac{e^{X\beta^i}}{\sum_{j=1}^{N} e^{X\beta^j}} \qquad \text{A.7}$$

In specification A.7, X denotes a vector of potential explanatory variables and β the vector of corresponding coefficients. One category is considered as the base mode (in this case the first category (no support) and the probabilities of the other categories are expressed relative to the base mode. The coefficient estimates then indicate whether an increase in the corresponding explanatory variable raises or decreases the probability that a company will be in a given mode (for example, receives a subsidy as well as a tax benefit) relative to the base mode (no support). The estimation of the selection model provides indications of the factors that appear to play a role in the (self-) selection of companies in applying for and receiving public support for their R&D activities but also provides the inverse Mills ratios. An inverse Mills ratio decreases monotonically as the probability that a company is selected (in effect, receives a subsidy; tax benefit or both) rises. A traditional Heckman model consists in a first-step Logit estimation with a binary choice (support vs. no support), resulting in one inverse Mills ratio. A Heckman model with a first step multinomial logit estimation results in N-1 inverse Mills ratios. These ratios are included in the second step estimation, the estimation of specification (1). Heckman (1979) noted that although the coefficients in the second step estimation are consistent, standard errors will be underestimated and, therefore, statistical significance overestimated. Maximum Likelihood (ML) estimation is needed to obtain correct standard errors. This requires a specification of the joint density function of the errors in the first-step selection equation and the second-step equation. As ML is an iterative process, it does not always converge. Especially in large samples, the two-step Heckman procedure is often preferred to the ML estimation (for example, Hussinger 2008).

The Heckman selection model assumes a joint normal distribution of the error terms in the selection specification and the (second-step) specification of interest. In several studies on the impact of R&D subsidies, matching procedures are used to accommodate the potential bias due to the (self-) selection (for example, Czarnitzki and Fier, 2002; Almus and Czarnitzki, 2003; Czarnitzki and Lopes-Bento, 2014). The advantage of this approach is that it is non-parametric and therefore does not impose any parametric assumptions. The basic idea of matching is to construct a control group of individuals which are as similar as possible with individuals from the group of individuals that receive a given treatment (treatment group) except for the fact that they do not receive the treatment. If all relevant heterogeneity between individuals that may determine the probability of receiving a treatment is taken into consideration – in effect, the conditional independence assumption (CIA) holds – the selection bias is removed and an unbiased estimate of the average treatment effect for firms that receive support can be obtained (see specification A.2).

In contrast with a panel regression, not all firms are considered in matching but only firms that receive support and firms that do not receive support but can be matched with firms that receive support. If a firm cannot be matched with a firm that does not receive support, it is also not considered in the matching estimation.

If many covariates need to be accounted for, matching can become problematic (curse of dimension). Rosenbaum and Rubin (1983) therefore proposed a balancing score b(X), which is a function of the covariates X that results in the conditional distribution of X, given that the balancing score is independent of whether an individual receives treatment or not. Most matching procedures use the propensity score as balancing score, which denotes the probability of an individual (firm) to receive treatment (support), conditional on observed covariates X (Caliendo and Kopeinig 2008: p. 36):

$$P_i(X) = \ P(D_i = 1|\ X) \qquad \text{A.8}$$

Under the conditional independence assumption and assuming overlap between the treatment and the control group[48], the Propensity Score Matching (PSM) estimator for the average treatment effect on the treated (ATT) is (Caliendo and Kopeinig 2008: p. 36):

$$\tau_{ATT} = \ E_{P(X)\,|\,D=1}\{E[Y(1)|D=1, P(X) - E[Y(0)|D=0, P(X)]]\} \qquad \text{A.9}$$

Different algorithms can be used to match individuals. Caliendo and Kopeinig (2008) distinguish four alternatives: Nearest neighbour; Caliper/radius; Stratification/interval and Kernel/Local Linear. In this paper the most straightforward algorithm, nearest neighbour matching is considered, which consists in matching a firm that receives support with a firm that does not receive support but is closest in terms of the propensity score. In the matching reported in section 3.8, replacement is allowed for as this increases the quality of matching and reduces the bias (at the expense of lower efficiency). Replacement implies that a firm that receives no support can be considered as a match for several firms with support. To avoid bad matches, a caliper is applied which imposes a maximum propensity score distance that is tolerated for a match (see, for example, Czarnitzki and Lopes-Bento (2013) and Hottenrott and Lopes-Bento (2014). More details on matching are provided in the survey by Caliendo and Kopeinig (2008). As panel data are used instead of cross-section data, exact matching on the year is imposed.

Instrumental variables

A similar problem to the (self-) selection bias is endogeneity (simultaneity) of public support and R&D expenditures. As pointed out by Jaffe (2002), if the amount of public support that a company receives – and not just the fact whether a company gets support or not – is related to unobservable variables or the autonomous decision of a company how much it wants to invest in R&D, the amount of public support is endogenous. Kauko (1996) argued that studies that ignore endogeneity tend to overestimate the effectiveness of R&D subsidies whereas studies in which endogeneity is accounted for indicate that subsidies are rather inefficient in stimulating private R&D activities. Although the problem of endogeneity in assessing the impact of public support for R&D cannot be denied, the solution is less straightforward. If a supposedly independent (exogenous) variable is in fact correlated with the error term, estimates based on the assumption of no correlation will not be consistent. Instrumental variables (IV) estimation can provide consistent estimates through the use of instruments, which are variables that correlate well with a potentially endogenous variable but do not correlate with the error term. The consistency of IV estimation rests on the assumption that valid instruments exist, in effect variables that

[48] The overlap or common support condition states that the probability of treatment should lie between 0 and 1 but 0 and 1 not included. In other words, the probability of an individual in the treatment (control) group to belong to the treatment group should not equal 1 (0).

satisfy both conditions. Cameron and Trivedi (2010) point out that this is a very strong assumption and that it is very difficult to find valid instruments. When instruments are not correlated with the error term, they may also not, or only weakly, be correlated with the variables they are supposed to instrument. In small samples, IV estimates are biased (Nelson and Startz 1990; Bound et al 1995; Angrist and Pishke 2009). Nelson and Startz (1988) argued that when instruments are weak, causal inference will be spurious and IV estimation may even perform worse than OLS. Weak instruments also result in less efficient estimates, i.e. large standard errors so that the statistical significance of coefficients is underestimated. Although tests exist to test instrument validity under certain conditions, Cameron and Trivedi (2010: p. 181) point out that instrument validity relies on persuasive argument, economic theory and previous empirical work rather than on rigorous testing. The most common IV procedure is two-stage least squares (2SLS). In a first stage, the variable that is assumed to be endogenous, X, is regressed on a potential instrument, Z (e.g. Angrist and Pishke, 2015: p. 132):

$$X_{it} = \alpha + \beta Z_{it} + \varepsilon_{it} \qquad \text{A.10}$$

The fit for X, i.e. $\widehat{X_{it}} = \hat{\alpha} + \hat{\beta} Z_{it}$, derived from this first step is then included in a second-step (structural):

$$Y_{it} = \gamma + \delta_{2SLS}\widehat{X_{it}} + \mu_{it} \qquad \text{A.11}$$

Both equations can include control variables. If the instrument Z is valid and relevant, δ_{2SLS} provides an unbiased estimate of the "causal" effect of X on Y. An instrument is valid if it is not correlated with the residual in the structural equation: Corr (Z_{it}, μ_{it}) = 0 and relevant if it is correlated with the endogenous variable it instruments: Corr (Z_{it}, X_{it}) ≠ 0. To test for instrument validity, the specification needs to be over-identified, which implies that more than one instrument is required for each endogenous variable. Under over-identification, tests on instrument validity can be performed. These tests are carried out in a Generalized Method of Moments (GMM) framework, which offers a generalization of IV and many other estimators (see next section).

Although IV estimation is the most common framework to tackle possible endogeneity of right-hand-side variables in regression, it should not be considered as the panacea for establishing causal effects. Hall and van Reenen (2000) point out that instrumental variables estimation results in the loss of precision (efficiency) in estimation and problems with finding appropriate instruments to identify the endogenous variables. Nichols (2007) enlists some problems of IV tests. Crown et al. (2011) compared the results of OLS to results from IV estimation, in simulations that account for varying degrees of endogeneity of the independent variable, weak and contaminated instruments and sample size. They conclude that although IV estimation is valuable for testing in the presence of endogeneity that only under the most ideal circumstances IV will produce estimates with less estimation error than OLS.

Imbens (2014) provides a recent survey on instrumental variables estimation in econometric studies.

Generalized Method of Moments

Fixed effects (within) estimation of panel data permits to account for differences between firms (unobserved heterogeneity) that do not change much over time. Within estimation however tends to magnify measurement error in explanatory variables, often resulting in low statistical significance and unreliable coefficient estimates (Griliches and Hausman 1986). As most firm-level data are known to be prone to measurement error, within estimation may lead to an underestimation of statistical significance. An alternative to within estimation, which considers deviations of the variables for each firm around its mean over time, is taking first differences, which eliminates time-invariant fixed effects. As pointed out by Griliches and Hausman (1986), measurement error likely results in a downward bias of first difference estimators that exceeds the bias of within estimators. However, together, a within and a first difference estimator provide an estimate of the measurement error bias and thus also a way to construct a bias-corrected estimator. Griliches and Hausman (1986) proposed to use additional estimators of the bias, based on differences over longer time periods (for example the difference between period t and t-2 or t-3). These estimators can be seen as instrumental variable estimators and the combination of the estimators intends to increase efficiency. The different moment conditions can be combined in a system of equations. Three-stage Least Squares (3SLS)[49] or Generalized Least Squares (GLS) estimators will not be consistent and for optimal estimation Griliches and Hausman (1986) used the Generalized Method of Moments (GMM) estimator. Most estimators are based on moment conditions. Ordinary Least Squares (OLS), for example, is based on the condition (assumption) that the explanatory variables x are not correlated with the residuals e, $E[x_i e_i] = 0$. Instrumental variable estimators are based on the assumption that instruments z are not correlated with the residuals, $E[z_i e_i]$. Most estimators based on moment conditions apply to exact identification, as the number of moment conditions that are used equals the number of parameters that need to be estimated. GMM intentionally considers more moment conditions than necessary, which results in over-identification. This may appear problematic as it cannot ensure a single solution. GMM is developed to reconcile possible conflicting estimates (Greene 2000, p. 479). In the context of instrumental variables estimation, the main advantage of GMM is that it allows to test a crucial assumption of instrument validity, in effect, that the instruments should not be correlated with the results. This test is not possible under exact identification.

Dynamic panel models are most often estimated with GMM. A dynamic panel is a specification in which persistence in the dependent variable is modelled by including one or more of its lags, for example a model with p lags of the dependent y, a matrix of explanatory variables x and individual (firm) time-fixed effects α_i (Cameron and Trivedi 2010: p. 293):

$$y_{it} = \gamma_1 y_{i,t-1} + \cdots + \gamma_p y_{i,t-p} + x'_{i,t}\beta + \alpha_i + \varepsilon_{i,t} \qquad \text{A.12}$$

As pointed out by Cameron and Trivedi (2010) the correlation in the dependent variable can result from different mechanisms:

- True state dependence: through direct effect of values of y in previous years
- Observed heterogeneity: through effects of explanatory variables x
- Unobserved heterogeneity: through time-fixed individual effects α_i

---

49 3SLS is a generalization of two-stage least squares (2SLS) that increases efficiency.

Cameron and Trivedi (2010) point out that the specific mechanism that explains the correlation in the dependent variable over time has policy relevance. If, in a given year, a large shock in $\varepsilon_{i,t}$ occurs, the shock will have a lasting effect if the coefficient of the lagged dependent, γ_1, is close to 1 but only a transitory effect if the coefficient is close to 0. A fixed effects (within) specification with lagged dependent variables will provide inconsistent estimates in short panels (many individuals over a short period of time).[50] A straightforward way to avoid this consists in taking first differences, which removes time-fixed effects (Cameron and Trivedi 2010: p. 294):

$$\Delta y_{it} = \gamma_1 \Delta y_{i,t-1} + \cdots + \gamma_p \Delta y_{i,t-p} + \Delta x'_{i,t}\beta + \Delta\varepsilon_{i,t} \qquad \text{A.13}$$

As $\Delta y_{i,t-1}$ is correlated with $\Delta\varepsilon_{i,t}$, OLS of the first-difference specification would provide inconsistent results, even if – as is assumed – $\varepsilon_{i,t}$ are not serially correlated. For consistent estimation instruments for the lagged first differences of the dependent variable are required and for the first differences of the explanatory variables if these are endogenous. Anderson and Hsiao (1981); Holtz-Eakin et al. (1988) and Arellano and Bond (1991) proposed to use lagged levels of y as instruments for the first-differences of y, for example $y_{i,t-2}$ as an instrument for $y_{i,t-1}$. As there are more instruments than coefficient to estimate, the specification is over-identified and GMM can be applied to test the validity and relevance of the instruments. Estimators of a first-difference (FD) dynamic panel with lagged levels as instruments is generally referred to as the FD-GMM or Arellano Bond estimators. FD-GMM can be estimated in one step or two steps, with the latter in principle providing more efficient estimation (lower standard errors). Some studies have found that the two-step GMM can have a large bias in small samples (Drukker 2008). Windmeijer (2005) proposed a finite-sample correction of the standard errors, which is applied in this paper.

Instruments for endogenous explanatory variables can be included in the GMM specification, either GMM-style instruments (lags) or own instruments.

For FD-GMM estimation to be consistent, $\varepsilon_{i,t}$ should not be serially correlated. Arellano and Bond (1991) proposed tests of serial correlation in a GMM specification. Absence of serial correlation would be indicated by the rejection of the null hypothesis of no serial correlation of order 1 of the first-differenced errors but no rejection of serial correlation at order 2 or more. (Cameron and Trivedi 2010: p. 300).

Arellano and Bover (1995) and Blundell and Bond (1998) have shown that under persistent serial correlation, lagged levels of the dependent variable can be weak instruments. Using the moment condition $E(\Delta y_{i,t-1}\varepsilon_{i,t}) = 0$ they proposed to consider a level specification, in addition to the first-difference specification, and use first differences as instruments for the level of the dependent variable. Using additional moment conditions, this System GMM is considered to more efficient estimates with better finite sample properties than the FD-GMM (Cameron and Trivedi 2010; Roodman 2009).

First differences magnify the gaps in an unbalanced panel. As shown in Table A.2.2 in Annex 2, the panel used in this paper is highly unbalanced. Arellano and Bover (1995) proposed “forward orthogonal deviations”, which consists in subtracting all available future observations from the contemporaneous

[50] In a long panel (observations over many years), the dynamic panel bias becomes insignificant so that a fixed effects specification provide reliable estimates (Roodman 2009).

value of a variable. This minimizes the loss of observations due to gaps and provides valid instruments as no lagged observations are considered (Roodman 2009).

In a GMM framework, an over-identified specification – more instruments than parameters to estimate – permits to test the validity of instruments. As mentioned before, Arellano and Bond (1991) proposed a test of serial correlation in the error term of a first-difference specification. Rejection of the null hypothesis at order one and no rejection at higher order serial correlation would indicate that the model is not wrongly specified. Rejection at higher order than 1 implies that the instruments (moment conditions) are not valid. Arellano and Bond (1991) also proposed a Sargan test of over-identifying restrictions. The test is known to over-reject the null hypothesis of joint validity of instruments if the error term is heteroskedastic. An alternative test of instrument validity is the Hansen test, which is more robust than the Sargan test but can be weak if there are many instruments. The Stata procedure XTABOND2, which is used to perform the GMM estimations in this paper, reports both tests.

XTABOND2 also reports a number of tests of the relevance of instruments, which consider the correlation between instruments and the endogenous variables they are supposed to instrument. As the null hypothesis is zero correlation, failure to reject these tests would indicate the irrelevance (weakness) of the instruments considered.

The use of GMM in empirical studies has been prolific although – as with all estimation procedures – GMM has its limitations. Roodman (2007, 2009) points out that GMM estimators easily generate numerous instruments, which especially in System GMM may even be suspect and provides implausible results for the test of the validity of instruments.

Error Correction Model (ECM)

A regression in which time series variables are not stationary may result in "spurious correlation", in effect, the conclusion that there is a real relationship between variables whereas the association is simply due to the fact that the variables are similarly trended. As many economic variables are not stationary (even in real terms) it is important to establish whether there is an actual link between the variables. This is the subject of cointegration analysis. A first step is to establish the degree of integration of the variables, which reflects the number of times a variable needs to be differenced to be stationary. If a variable is stationary in levels, it is integrated of order 0, denoted by I (0). If it needs to be differenced n time before it is stationary it is integrated of order n, I (n). The order of integration of a variable is established through unit root tests. Dickey and Fuller (1979) introduced a unit root tests which consists in the OLS estimation of the following specification:

$$y_t = \rho y_t + u_t \qquad u_t \sim N(0, \sigma^2) \tag{A.14}$$

If $\rho = 1$, variable y contains a unit root and is not stationary. A constant can be included which, if statistically significant would indicate a variable with unit root and drift. If a trend variable is included and its coefficient is significant, the variable contains a unit root with trend.

An augmented Dickey-Fuller test accounts for serial correlation by considering the following specification:

$$\Delta y_t = \beta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + \cdots + \varepsilon_t \qquad \text{A.15}$$

The augmented Dickey-Fuller test consists in testing the null hypothesis: $\beta = 0$.

The Dickey-Fuller test was shown to have low statistical power, in effect the unit root null hypothesis (no stationarity) hypothesis is too often not rejected even if the variable is stationary. Other unit root tests have been developed but these tests can generally not distinguish stationary processes that are highly persistent from non-stationary processes (e.g. Zivot and Jiahui 2002).

Unit root tests were originally developed for single time series. A panel contains, for a given variable, time series for each individual. Panel data may increase the power of unit root tests because of the cross-section but also pose specific problems because of the need to account for unobserved cross-section heterogeneity (Cameron and Trivedi 2010: p. 279). Several panel unit root tests exist that account for panel heterogeneity and other assumptions on serial correlation and trends. In this paper the Stata procedure XTFISHER is used which, in contrast with most other tests, allows for unbalanced panels. XTFISHER is based on the p-values of individual unit root tests. The null hypothesis of the Fisher test is that all time series are not stationary against the alternative that at least one time series is stationary.

An error-correction model (ECM) attempts to separate short-run from long-run effects. An ECM specification can be derived by inserting a long-term specification into an autoregressive distributed lags (ADL) equation. Using the baseline specification and considering a one lag ADL results in[51]:

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \textstyle\sum_{s=1}^{S} \gamma_s \, \Delta x_{i,t} + \varphi\left(y_{i,t-1} - \sum_{s=1}^{S} \tau_s \, x_{i,t-1}\right) + \varepsilon_{i,t} \qquad \text{A.16}$$

Variable y is the dependent variable (log of own R&D expenditures) and the variables x are the support variables (also in logs). The coefficients of the first difference variables, γ_s, reflect the short-run effects of public support on private R&D expenditures. The level specification reflects the long-run cointegration relationship between support and R&D expenditures. The error correction coefficient φ should lie between -1 and 0 as this implies that any deviation from the long-run relationship is corrected, in effects entails a return towards the long-run relationship. For a long-run (cointegration) relationship between R&D expenditures and public support to exist, the variables need to be cointegrated, which implies that they should have the same order of integration and the error term should be stationary. As shown in section 3.9 the Fisher panel unit root test appears to cast doubt on both assumptions. It should be stressed that the panel in this paper is rather short, for which concepts as stationarity and cointegration are not very meaningful. The results of the ECM, as well as the results on the dynamic panel estimation, in this paper, should therefore be considered as a first attempt to account for the dynamic process and the high persistence in the variables. However, given the evidence of strong persistence in R&D expenditures, it seems important to consider a dynamic specification in future evaluations that can cover a longer period.

51 The ECM specification follows the derivation in a rather similar context by Mulkay and Mairesse (2013). The control variables are not shown but included in the estimation.

Annex 2: Descriptive statistics

Graph A.1 Distribution (kernel) of the log of R&D expenditures (2011)

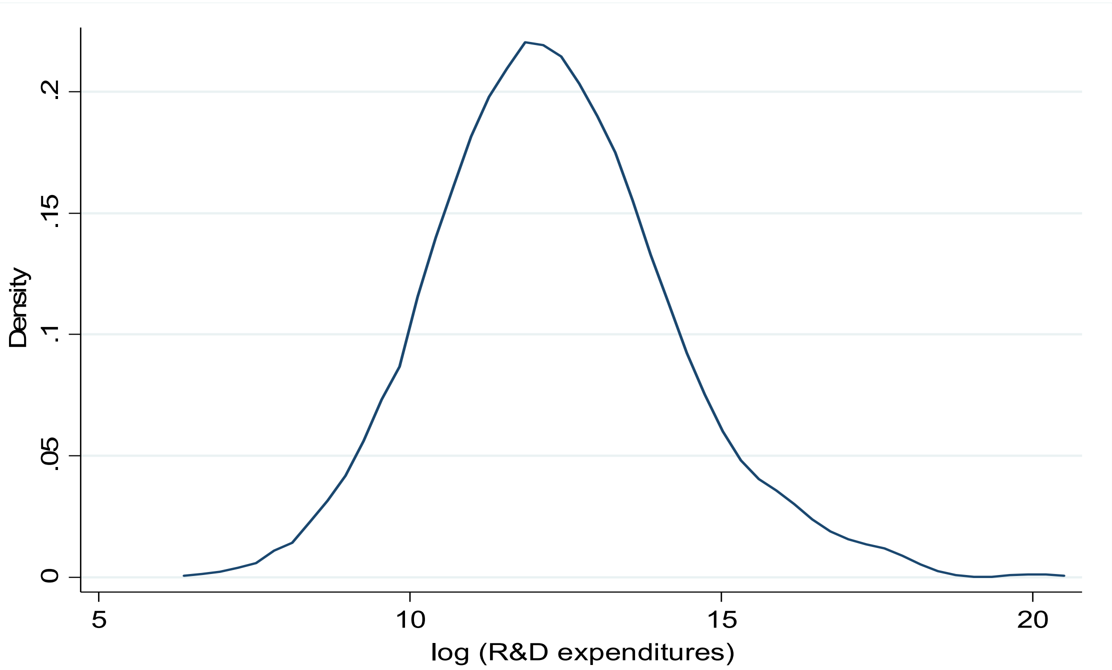


Table A.2.1 Descriptive (standardized) statistics by support scheme (2011)

	No support		Subsidy		Exemption cooperation		Exemption YIC	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
R&D expenditures	-0.12	-0.21	0.42	-0.19	0.73	-0.13	-0.13	-0.19
R&D/value added	-0.12	-0.23	0.40	-0.24	0.73	-0.12	-0.11	-0.22
Total public support	-0.23	-0.22	1.09	0.14	0.98	0.00	0.16	-0.06
# employees (FTE)	-0.09	-0.31	0.26	-0.25	0.49	-0.18	-0.30	-0.32
Firm age	0.02	-0.18	-0.09	-0.37	0.21	-0.09	-0.69	-0.75
Cash flow	-0.07	-0.18	0.20	-0.15	0.35	-0.12	-0.24	-0.16
Capital/employee	-0.02	-0.25	0.03	-0.22	-0.03	-0.27	-0.10	-0.27

	Exemption PhD		Exemption master		Tax credit R&D		Tax deduction patent income	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
R&D expenditures	0.67	-0.05	0.63	-0.08	1.51	-0.03	1.30	-0.07
R&D/value added	0.73	0.04	0.70	-0.03	1.41	0.08	1.35	-0.02
Total public support	0.78	0.00	0.80	-0.01	1.97	0.27	1.53	0.27
# employees (FTE)	0.56	-0.02	0.56	-0.01	1.08	-0.02	0.66	-0.06
Firm age	0.21	0.03	0.15	0.01	0.12	-0.45	0.18	-0.11
Cash flow	0.42	-0.06	0.41	-0.06	0.96	-0.08	1.12	0.07
Capital/employee	0.05	-0.22	0.00	-0.25	-0.10	-0.29	0.20	-0.08

Note: This table shows descriptive statistics for firm-level variables that have been standardized with regard to the average in the industry to which a firm belongs:

$$Standardized\ X = \frac{(x - industry\ average(x))}{industry\ standard\ deviation\ (X)}$$

Table A2.2: Data availability 2003-2011

Data availability	Frequency	%	Cumulative %
.......11	578	18.22	18.22
.....1111	269	8.48	26.7
.....11..	187	5.9	32.6
1........	143	4.51	37.11
...111111	135	4.26	41.36
...1111..	132	4.16	45.52
.11......	128	4.04	49.56
111111111	123	3.88	53.44
...11....	107	3.37	56.81
Other pattern	1370	43.19	100
Total number of firms	3172		

Note: This table shows the time pattern of observations occurring in the panel. The number ***1*** denotes the fact that for a given year there are no missing values for a firm whereas a point(.) denotes that a firm has missing values, for at least one variable in the baseline specification, such that for that year the firm is not considered in the estimation.

Annex 3: Results by industry

Table A3.1 Number of observations by NACE industry

NACE (rev.2) two-digit	Description	# observations
1	Crop and animal production, hunting and related service activities	102
2	Forestry and logging	47
3	Fishing and aquaculture	21
6	Extraction of crude petroleum and natural gas	12
7	Mining of metal ores	28
8	Other mining and quarrying	87
10	Manufacture of food products	1246
11	Manufacture of beverages	118
12	Manufacture of tobacco products	43
13	Manufacture of textiles	546
14	Manufacture of wearing apparel	125
15	Manufacture of leather and related products	31
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	197
17	Manufacture of paper and paper products	196
18	Printing and reproduction of recorded media	313
19	Manufacture of coke and refined petroleum products	16
20	Manufacture of chemicals and chemical products	926
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	211
22	Manufacture of rubber and plastic products	698
23	Manufacture of other non-metallic mineral products	521
24	Manufacture of basic metals	251
25	Manufacture of fabricated metal products, except machinery and equipment	1166
26	Manufacture of computer, electronic and optical products	514
27	Manufacture of electrical equipment	381
28	Manufacture of machinery and equipment n.e.c.	961
29	Manufacture of motor vehicles, trailers and semi-trailers	324
30	Manufacture of other transport equipment	74
31	Manufacture of furniture	272
32	Other manufacturing	247
33	Repair and installation of machinery and equipment	271
35	Electricity, gas, steam and air conditioning supply	82
37	Sewerage	47
38	Waste collection, treatment and disposal activities; materials recovery	252
39	Remediation activities and other waste management services	36
41	Construction of buildings	401
42	Civil engineering	264
43	Specialised construction activities	882
45	Wholesale and retail trade and repair of motor vehicles and motorcycles	443
46	Wholesale trade, except of motor vehicles and motorcycles	3131
47	Retail trade, except of motor vehicles and motorcycles	220

NACE (rev.2) two-digit	Description	# observations
49	Land transport and transport via pipelines	607
50	Water transport	12
51	Air transport	11
52	Warehousing and support activities for transportation	397
53	Postal and courier activities	45
55	Accommodation	14
56	Food and beverage service activities	38
58	Publishing activities	248
59	Motion picture, video and television programme production, sound recording and music publishing activities	113
60	Programming and broadcasting activities	42
61	Telecommunications	173
62	Computer programming, consultancy and related activities	1995
63	Information service activities	112
64	Financial service activities, except insurance and pension funding	272
65	Insurance, reinsurance and pension funding, except compulsory social security	78
66	Activities auxiliary to financial services and insurance activities	223
68	Real estate activities	34
69	Legal and accounting activities	217
70	Activities of head offices; management consultancy activities	614
71	Architectural and engineering activities; technical testing and analysis	1561
72	Scientific research and development	543
73	Advertising and market research	181
74	Other professional, scientific and technical activities	132
77	Rental and leasing activities	73
78	Employment activities	247
79	Travel agency, tour operator and other reservation service and related activities	75
80	Security and investigation activities	79
81	Services to buildings and landscape activities	280
82	Office administrative, office support and other business support activities	179
85	Education	40
86	Human health activities	100
88	Social work activities without accommodation	38
95	Repair of computers and personal and household goods	10
96	Other personal service activities	15

Table A3.2 **Results of panel estimation by industry (2003-2011) - Top 10 industries by public support**

	Computer programming, consultancy and related activities	Architectural and engineering activities; technical testing and analysis	Wholesale trade, except of motor vehicles and motorcycles	Manufacture of chemicals and chemical products
Dependent variable (R&D expenditures net of public support)				
Explanatory variables:				
Regional subsidy	0.05 (1.29)	0.09 (1.73)*	0.25 (2.78)***	0.03 (1.14)
Research cooperation	-0.02 (-0.40)	0.29 (1.68)*	0.16 (1.03)	0.12 (1.10)
Young Innovative Company	-0.03 (-0.50)	-0.04 (-0.37)	-0.08 (-1.57)	-0.00 (-0.04)
PhDs and civil engineers	0.01 (0.20)	0.06 (0.78)	0.13 (1.84)*	0.01 (0.13)
Master	0.19 (2.41)**	0.13 (1.69)*	0.21 (2.33)**	-0.00 (-0.05)
Tax credit R&D	-0.07 (-1.34)	0.09 (1.44)	0.47 (1.40)	-0.11 (-2.86)***
Tax deduction 80% patent income	-0.03 (-0.41)	-0.03 (-0.48)	-0.07 (-0.75)	-0.01 (-0.09)
Control variables:				
Cash flow	0.37 (1.40)	0.15 (0.56)	0.45 (1.45)	-0.24 (-1.40)
Number of employees	1.87 (2.69)***	2.20 (2.60)***	1.97 (1.62)*	1.03 (1.51)
Age	-0.40 (-1.17)	-0.14 (-0.37)	-0.86 (-1.75)*	0.59 (1.63)*
Capital intensity	0.37 (1.01)	0.64 (1.48)	0.31 (0.85)	-0.01 (-0.01)
Number of observations	919	683	782	608

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification for individual industries. All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity.

Table A3.2 (continued) **Results of panel estimation by industry (2003-2011) - Top 10 industries by public support**

	Manufacture of machinery and equipment n.e.c.	Scientific research and development	Activities of head offices; management consultancy activities	Manufacture of computer, electronic and optical products
Dependent variable (R&D expenditures net of public support)				
Explanatory variables:				
Regional subsidy	0.08 (2.14)**	-0.01 (-0.28)	0.04 (0.33)	0.01 (0.72)
Research cooperation	0.02 (0.51)	0.07 (1.00)	0.49 (1.22)	-0.03 (-0.89)
Young Innovative Company	8.10 (1.67)*	-0.02 (0.16)	-1.01 (-2.67)***	0.01 (0.11)
PhDs and civil engineers	0.00 (0.06)	0.09 (0.72)	0.04 (0.19)	-0.02 (-0.60)
Master	-0.01 (-0.26)	0.12 (2.51)***	0.61 (2.84)***	0.03 (0.89)
Tax credit R&D	0.02 (0.54)	0.05 (1.03)	0.38 (2.31)**	-0.04 (-0.93)
Tax deduction 80% patent income	0.00 (-0.17)	-0.03 (-0.43)	0.13 (0.30)	0.05 (1.92)*
Control variables:				
Cash flow	-0.06 (-0.22)	-0.19 (-0.95)	-0.35 (-0.51)	0.23 (1.25)
Number of employees	0.72 (0.99)	0.99 (2.69)***	3.79 (3.46)***	-0.25 (-0.30)
Age	-0.16 (-0.60)	0.03 (0.06)	-0.64 (-1.64)*	-0.10 (-0.53)
Capital intensity	0.11 (0.31)	0.23 (1.68)*	-0.69 (-0.98)	0.16 (0.87)
Number of observations	592	209	137	345

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification for individual industries. All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity

Table A3.2 (continued) **Results of panel estimation by industry (2003-2011) - Top 10 industries by public support**

	Manufacture of food products	Manufacture of fabricated metal products, except machinery and equipment
Dependent variable (R&D expenditures net of public support)		
Explanatory variables:		
Regional subsidy	0.02 (0.33)	0.05 (0.72)
Research cooperation	-0.05 (-0.73)	0.16 (2.40)**
Young Innovative Company	-	-
PhDs and civil engineers	0.04 (0.59)	0.05 (0.34)
Master	0.08 (0.93)	0.37 (2.38)**
Tax credit R&D	-1.57 (-0.88)	0.26 (1.91)*
Tax deduction 80% patent income	0.01 (0.25)	-0.03 (-0.48)
Control variables:		
Cash flow	0.42 (1.10)	-0.10 (-0.29)
Number of employees	0.02 (0.01)	1.14 (0.60)
Age	0.19 (0.37)	-0.35 (-0.63)
Capital intensity	-0.03 (-0.04)	0.20 (0.25)
Number of observations	583	419

Note: The table shows the results of a fixed effects (within) estimation of the baseline panel specification for individual industries. All variables are considered in logs, except firm age.

*, ** and *** denotes that the coefficient estimate differs from zero at a statistical significance level of respectively 10%, 5% and 1%. The t-values, shown in brackets, are robust to heteroskedasticity