École des Hautes Études en Sciences Sociales

Doctoral School n°465 Économie Panthéon Sorbonne

Paris-Jourdan Sciences Économiques

PHD THESIS

Discipline: Economics

Matthieu Lequien

Essays in economics

Innovation, firm dynamics and financial cycles

Thesis supervised by **Philippe Aghion** (Collège de France)

Date of defense: December 9^{th} 2021

Referees: Thierry Mayer (Sciences Po)
Isabelle Méjean (Sciences Po)

Jury: Philippe Aghion (Collège de France)
Roger Guesnerie (Collège de France)
Xavier Jaravel (LSE)

École des Hautes Études en Sciences Sociales

École doctorale n°465 Économie Panthéon Sorbonne

Paris-Jourdan Sciences Économiques

DOCTORAT

Discipline: Analyse et Politique Économiques

Matthieu Lequien

Essais en économie

Innovation, dynamique d'entreprises et cycles financiers

Thèse dirigée par **Philippe Aghion** (Collège de France)

Date de la soutenance: 9 décembre 2021

Rapporteurs: Thierry Mayer (Sciences Po)
Isabelle Méjean (Sciences Po)

Jury: Philippe Aghion (Collège de France)
Roger Guesnerie (Collège de France)
Xavier Jaravel (LSE)

Remerciements

Je remercie toutes les personnes qui m'ont aidé tout au long de cette thèse, d'une manière ou d'une autre. J'ai appris énormément des interactions que j'ai eues avec de nombreuses personnes, et cela a contribué à faire de moi un meilleur économiste.

Le premier de cette liste est évidemment l'homme qui a accepté de me guider et de diriger ma thèse, Philippe Aghion. A tout moment, il a trouvé le temps de m'aider et de m'écouter quand j'en avais besoin. J'apprécie particulièrement son attitude toujours positive, se réjouissant sincèrement quand les projets se déroulent bien, et très encourageant quand ils sont sur le point de bien se dérouler. Son énergie et son enthousiasme incroyables sont communicatifs et constituent un merveilleux carburant pour mener des recherches fructueuses. Malgré son agenda rempli de tonnes de projets, il parvient à rester proche de ses co-auteurs et de ses doctorants, et désireux d'aider dès qu'il le peut. J'ai évidemment beaucoup appris à son contact, bien sûr en économie, mais aussi plus généralement sur la manière de mener ses recherches, depuis l'émergence de l'idée à la publication.

Je tiens aussi à remercier tous les co-auteurs des chapitres de cette thèse, de qui j'ai tant appris: Antonin Bergeaud, qui a partagé mon bureau plusieurs années et avec qui j'ai la chance de travailler sur trois articles; Gábor Kátay, qui a aussi partagé mon bureau plusieurs années, pour les innombrables discussions, toujours très stimulantes, que nous avons pu avoir, sur l'économie et la vie en général; Lisa Kerdelhué, dont j'apprécie particulièrement la bienveillance et la bonne humeur; Marc J Melitz, avec qui j'ai aussi la chance de co-écrire trois articles, et qui m'impressionne par l'intelligence de ses raisonnements et son extrême gentillesse; et Stefanie Stantcheva, de laquelle j'ai beaucoup appris en termes de méthodes de travail notamment.

Je remercie ensuite tous mes collègues au cours de ces années, qui ont rendu l'économie – et la vie – si passionnantes. En particulier Antoine Devulder, pour de nombreuses discussions sur la modélisation, extrêmement riches intellectuellement, et l'honnêteté des raisonnements; Martin Mugnier et Loriane Py, avec qui j'ai constitué la base de données de brevets utilisée dans le chapitre 2; Sébastien Roux et Olivier Simon, pour m'avoir fait confiance à l'Insee ainsi que pour la bienveillance et la richesse des échanges que nous avons pu avoir.

Je souhaite également remercier mes parents, pour leurs questionnements bienveillants sur le contenu de ma thèse et sur la date prévue pour la soutenir. Enfin, et surtout, je remercie ma femme pour son soutien indéfectible et ô combien précieux tout au long de ces années.

Summary

This thesis is composed of three independent chapters. Chapter 1 describes how the demand conditions in their export markets shape French firms' innovation decisions. Chapter 2 analyses how (small) French firms react to a set of regulations of the different fiscal regimes that they can file in. Chapter 3 explains how adding financial information to standard output gaps models can help inform on the cyclical state of 8 advanced economies.

The first chapter analyses how the evolution of a firm's export markets influences its innovation decisions. French manufacturing firms patent more when they are subject to a foreign demand shock, i.e. when their export markets exogenously grow more. The increase in the number of patents is visible two to five years after such a demand shock, highlighting the time needed to innovate. This effect is entirely attributable to the initially most productive firms. In addition to its effects on innovation, a foreign demand shock has a positive and immediate impact on the sales and employment of firms, regardless of their productivity. This skewed innovation response to common demand shocks arises naturally from a model of endogenous innovation and competition with firm heterogeneity. The market size increase drives all firms to innovate more by increasing the innovation rents; yet by inducing more entry and thus more competition, it also discourages innovation by low productivity firms.

The second chapter analyzes how French self-employed individuals respond to a notch in their tax schedule, and identifies the motives behind that response. French self-employed can choose between three fiscal regimes, each of which differs in terms of tax incentives and administrative simplicity. In the two simplest regimes, the self-employed show a massive bunching just below the turnover threshold above which the individuals no longer qualify for those regimes. Entrepreneurs respond to tax incentives, as individuals with higher marginal tax rates bunch more. Many behavioral patterns suggest that this observed bunching comes at least to some extent from misreporting: just below the threshold, individuals report more often round numbers and appear to shift income to their partner when both are self-employed. In addition, individuals appear to learn over time how the system works – and how to take advantage of it. Finally a structural model of regime choice, estimated by fitting data moments, confirms the role of tax evasion and shows that the self-employed place a high value

on tax simplicity, particularly for the super simplified regime.

The third chapter assesses how to take into account financial cycles in the estimation of the output gap for eight advanced economies using two unobservable components models – an extended Hodrick-Prescott filter and a semi-structural model – and a semi-structural vector autoregression model in which only supply shocks are identified. Adding these financial variables (business credit or housing prices) makes it possible to capture the influence of the financial system on the business cycle that would not translate in inflation, capacity utilization rate or unemployment, the usual indicators of the cycle. The quality of the estimate of the output gap is assessed through its performance in predicting recessions. Models with financial variables generally produce better estimates of the output gap at the expense of increased volatility in real time. While the extended Hodrick-Prescott filter appears particularly attractive for its real-time stability, its predictive performance is weaker than that of the two semi-structured models. The vector autoregression model augmented with financial variables has the best ex-post prediction performance and similar performance to the semi-structural unobservable components model in real time. Overall, financial cycles appear to be important in Japan, Spain, the United Kingdom and, to a lesser extent, the United States and France, while they are relatively muted in Canada, Germany and Italy.

Keywords: Innovation, export, demand shocks, patents; taxation, bunching, evasion, self-employment; unobserved components model, semi-structural VAR, output gap, financial cycle

Résumé

Cette thèse est constituée de trois chapitres indépendants. Le premier chapitre décrit comment l'évolution des marchés d'exportation d'une entreprise influe sur ses décisions d'innovation. Le deuxième chapitre analyse la façon dont les (petites) entreprises françaises réagissent à la réglementation fiscale régissant les trois régimes fiscaux auxquels elles peuvent prétendre. Le troisième chapitre détaille comment intégrer des informations sur les déséquilibres financiers dans différents modèles d'écart de production éclaire la position dans le cycle de 8 économies avancées.

Le premier chapitre analyse dans quelle mesure l'évolution des marchés d'exportation d'une entreprise influe sur ses brevets. Les entreprises manufacturières françaises brevettent davantage lorsqu'elles sont soumises à un choc de demande étrangère, c'est-à-dire lorsque leurs marchés d'exportation croissent plus fortement de manière exogène. La hausse du nombre de brevets est perceptible deux à cinq ans après un tel choc de demande, mettant en évidence le temps nécessaire pour innover. Cet effet est entièrement imputable aux entreprises initialement les plus productives. Outre ses effets sur l'innovation, un choc de demande étrangère a un effet positif et immédiat sur les ventes et l'emploi des entreprises, et ceci quel que soit leur niveau de productivité. Cette réaction asymétrique de l'innovation à de même chocs de demande découle naturellement d'un modèle d'innovation et de concurrence endogènes avec hétérogénéité des entreprises. La hausse de la taille du marché incite toutes les entreprises à innover davantage en augmentant les rentes liées à l'innovation ; cependant, en induisant plus d'entrées et donc plus de concurrence, elle décourage également l'innovation des entreprises à faible productivité.

Le deuxième chapitre analyse la façon dont les entrepreneurs individuels français réagissent à un saut dans leur imposition, et en identifie les ressorts. Les indépendants français peuvent choisir entre trois régimes fiscaux, qui diffèrent chacun en termes d'incitations fiscales et de simplicité administrative. Dans les deux régimes les plus simples, beaucoup d'indépendants déclarent un chiffre d'affaires juste en dessous du seuil au-delà duquel les individus ne peuvent plus bénéficier de ces régimes. Les entrepreneurs réagissent aux incitations fiscales: la masse d'individus en excès juste en dessous du seuil est plus importante pour ceux qui ont un taux marginal d'imposition plus élevé. Plusieurs schémas de comportements suggèrent que ce regroupement d'individus observé tout près du seuil provient

au moins pour partie de déclarations inexactes : juste en dessous du seuil, les individus déclarent plus souvent des chiffres ronds et semblent transférer le revenu à leur partenaire lorsque tous deux sont indépendants. En outre, les individus semblent apprendre avec le temps le fonctionnement du système fiscal – et comment en tirer parti. Enfin, un modèle structurel de choix de régime, estimé sur données fiscales, confirme le rôle de l'évasion fiscale et montre que les indépendants accordent une grande valeur à la simplicité fiscale, en particulier pour le régime super simplifié.

Le troisième chapitre évalue comment prendre en compte les cycles financiers dans l'estimation de l'écart de production pour huit économies avancées à l'aide de deux modèles à composantes inobservables – un filtre Hodrick-Prescott élargi et un modèle semi-structurel – et un modèle semi-structurel à vecteur autorégressif dans lequel seuls les chocs d'offre sont identifiés. Ajouter ces variables financières (crédit aux entreprises ou prix immobiliers) permet de capter l'influence de la sphère financière sur le cycle de l'économie qui ne se verrait pas sur l'inflation, le taux d'utilisation des capacités de production ou le chômage, les indicateurs usuels du cycle. La qualité de l'estimation de l'écart de production est évaluée à travers sa performance à prévoir les récessions. Les modèles avec variables financières produisent généralement des estimations de l'écart de production plus performantes au détriment d'une volatilité accrue en temps réel. Alors que le filtre Hodrick-Prescott élargi apparaît particulièrement attractif pour sa stabilité en temps réel, sa performance de prédiction est plus faible que celle des deux modèles semi-structurels. Le modèle à vecteur autorégressif augmenté avec des variables financières présente les meilleures performances de prévision ex-post et des performances similaires au modèle semi-structurel à composantes inobservables en temps réel. Dans l'ensemble, les cycles financiers apparaissent importants au Japon, en Espagne, au Royaume Uni et, dans une moindre mesure, aux États-Unis et en France, alors qu'ils sont relativement contenus au Canada, en Allemagne et en Italie.

Mots-clés: Innovation, exportations, chocs de demande, brevets ; fiscalité, masse en excès, évasion, travail indépendant ; modèle à composantes inobservables, VAR semi-structurel, écart de production, cycle financier

Contents

Chapter 1

General Introduction

This thesis is composed of three independent chapters. Chapter 1 describes how the demand conditions in their export markets shape French firms' innovation decisions. Chapter 2 analyses how (small) French firms react to a set of regulations of the different fiscal regimes that they can file in. These two chapters answer from a different perspective the question : how do firms react to their environment? Whether this environment is taken as the conditions of the markets they operate in or the regulatory framework, chapters 1 and 2 illustrate its profound impact on the firms' decisions. Chapter 3 sheds some light on this environment using a macro lens to analyse the cyclical position of the economy which firms live in. It details how adding financial information to standard output gaps models can help inform on the cyclical state of 8 advanced economies.

Chapter 1 highlights one channel through which the decisions of a firm are shaped by the world in which it operates. Innovation in a firm depends on its expected return, itself strongly related to the size of the market the firm is expected to cover. This chapter shows that the economic magnitude of this link between market size and innovation is substantial. In the more conservative specification, we find that a 1 percent expansion/contraction in export demand leads to 52 additional/fewer priority patents (corresponding to the first patent publication for an invention) in the French manufacturing sector – a .64 aggregate elasticity. We analyze how the quantity and quality of this innovation response unfold over time and varies across firms with different initial levels of productivity.

In order to analyze those patenting responses, we merge comprehensive patent records with exhaustive firm-level production and customs data, which cover the whole population of French manufacturing firms. The combined use of these datasets has been made possible by a new algorithm developed in Lequien et al. (2019) that matches a French firm's name with its unique identifier (*Siren*) used in all French administrative business records and allows us to link the innovation activities of a firm with the other firm data sources. Innovation is measured by the flow of *priority* patent applications. The subsequent filings of the same intellectual property (in particular if they are filed at patent authorities in other countries) are secondary filings. We focus on priority patents for two reasons. First because our goal is not to measure a response in patenting but a response in innovation. By focusing on priority patents, we concentrate on patents that correspond to new inventions. Second because we want to avoid capturing the fact that firms that are more involved in international trade are more likely to patent many secondary filings so as to protect their invention in their sales' destinations.

Chapter 1 sheds light on three sets of results. First, on average firms respond to a positive export demand shock by innovating more. Since our specifications always control for sector-year effects, this innovation response must be driven by differences in firm-level innovation responses to demand shocks within each sector. This stands in sharp contrast to the literature measuring sector-wide innovation responses – whether across sectors or for a given sector over time.[1]

Second, the innovation response to a positive export demand shock takes 2 to 5 years to materialize in priority patents filings, highlighting the time needed for innovation. In contrast, sales and employment adjust immediately to the export shock. We interpret this difference as a confirmation that the response to export demand shocks captures a market size effect.

Third, the impact of a positive export demand shock on innovation is entirely driven by French firms with above median productivity levels (in an initial period prior to the demand shocks). This heterogeneous response could simply reflect the fact that the demand shock only affects the most productive firms. We check that this is not the case by allowing for a different impact of the export demand shocks on sales or employment depending upon initial productivity levels. We find that in contrast to what we observe for innovation, there is no heterogeneous response of sales or employment to a demand shock for low versus high productivity firms. Thus, similar demand shocks only lead to future innovation responses by relatively more productive firms.

These results provide some additional context to the recent literature documenting the rise of superstar firms: the skewed innovation response is likely to generate further increases in market share for the best performing firms leading to increases in market concentration. Indeed, Autor et al. (2017) document that this growth in concentration is most apparent in industries with above average growth in patent-intensity.

Our identification strategy relies on the construction of firm-level demand shocks that respond to aggregate conditions in a firm's export destinations but are independent of firm-level decisions (including the concurrent decisions for export-market participation and the forward looking innovation response). Following Hummels et al. (2014), this type of export demand shock has been used extensively in the recent empirical trade literature. It leverages detailed information on the set of products exported to specific destinations by a firm at a prior given date (prior to any changes in innovation that we analyze in our sample). Focusing on this export-driven measure of market size means that we are abstracting from the potential effects of domestic market demand on firms' innovation. For

[1] In an influential study, Acemoglu and Linn (2004) measure the sector-wide innovation response of the pharmaceutical industry to changes in demand over time.

this market, we cannot separate out the causal effects of domestic market size on innovation from the reverse effect of innovation on domestic demand and market size.

We show that our results using this identification strategy are robust to many different specifications including variations in the measure of and functional forms for innovation. We also perform placebo tests that independently confirm that our causation inference from increases in market size to innovation are well founded.

While several explanations might be entertained to explain why the effect of export on innovation should be skewed towards more frontier firms, we show that this outcome arises naturally from a model of exports and innovation with endogenous innovation and markups. In this setting, a positive export demand shock induces not only a direct market size effect – which increases innovation for all firms – but also a competition effect. An increase in market size in any export destination will attract new firms into the export market as more firms find it profitable to sell there. And indeed we find a positive correlation between our export demand shocks and various measures of firm entry into the corresponding destination markets. With endogenous markups (linked to endogenous price elasticities), this competition effect associated with entry impinges disproportionately on the market share of the less productive firms, reducing their incentives to innovate. Overall, this combination of the direct market size effect and of its induced competition effect leads to a skewed innovation response between more and less productive firms. Firms closest to the technological frontier increase innovation the most, while the combined effect can even be negative for the least productive firms.

Chapter 1 relates to several strands of literature. There is first the theoretical literature on trade, innovation and growth (see Grossman and Helpman, 1991a,b, Aghion and Howitt, 2009, chapter 13, and more recently Akcigit et al., 2018).[2] Our paper also relates to the recent empirical literature on firm-level trade and innovation. In particular both Lileeva and Trefler (2010) and Bustos (2011) highlight a clear relationship between R&D efforts and export status. Our analysis contributes to this literature in two main respects: (i) this literature focuses on the *extensive* margin of export markets (i.e. whether a firm exports or not to a particular market or set of markets) whereas we consider instead the effect of the *intensive* margin of exports (i.e. of the size of export markets) on innovation; (ii) we use innovation outcomes - the flow of priority patent filings - instead of R&D spending as our main measure of innovation, whereas these papers consider the causal impact of new export markets on R&D spending.[3]

[2] Akcigit et al. (2018) develop and calibrate a new dynamic trade model where firms from different countries compete strategically for leadership in domestic import and export markets. Their model predicts that trade openness encourages innovation in advanced sectors and discourages it in backward sectors. Dhingra (2013) and Impullitti and Licandro (2018) also develop theoretical models with endogenous firm innovation and endogenous competition (via endogenous markups). Dhingra (2013) focuses on the firm-level trade-offs between innovation and product variety, whereas Impullitti and Licandro (2018) focuses on the consequences of innovation for growth and welfare.

[3] In related work, Coelli et al. (2016) document the patenting response of firms in response to the Uruguay round of tariff levels.

There is also a recent literature on trade and innovation that focuses on the impact of import competition on domestic firms (see Bloom et al., 2016; Iacovone et al., 2011; Autor et al., 2016; Bombardini et al., 2017). These papers investigate whether import competition induces firms to innovate more in order to escape competition as in Aghion et al. (2005). Empirically, chapter 1 is quite distinct as it covers the market expansion channel related to exports.

Finally, this first chapter contributes to the empirical literature on market size and innovation, starting with Acemoglu and Linn (2004). We add to this literature in three main respects: (i) by providing evidence of a widespread (manufacturing) firm-level market size effect that is not driven by any sector-level dynamics; (ii) by showing that this market size effect is skewed and mainly driven by the most productive French firms; (iii) by looking at the time dynamics of the market size effect of expanded export markets on firm-level innovation: in particular we show that while a positive export demand shock immediately increases the firm's sales, the innovation response takes several additional years to materialize in new patents. However, one should keep in mind that our analysis is grounded in the market size variations arising from export destinations, which means that we do not use variations coming from the domestic markets. Thus we leave open the question as to whether the domestic market size affects firms in a similar way as the export market size effect that we document.

While Chapter 1 describes how firms react to the economic conditions they encounter, Chapter 2 sheds light on how firms react to the regulatory framework they face. More precisely, it tries to disentangle the relevance of different mechanisms – tax simplicity or fiscal incentives – in explaining the choice of tax regime by self-employed individuals in France, in conjunction with the ability to evade taxation.

Designing a policy that fulfills its stated goals, provides clear and correct incentives without unintended consequences, minimizes administrative hassle for individuals, and at the same time remains sufficiently simple for people to understand is an enormous challenge. Tax policy is a case in point: the best tax incentives may turn out to be ineffective if people do not understand them. Even worse, complexity may make the system more regressive if it is mostly the least sophisticated agents or those who cannot afford professional tax advice who cannot understand it and benefit from it. In the setting that this chapter considers, self-employed individuals can choose between three different tax regimes, balancing considerations on the simplicity, the fiscal incentives, and the ease of evading taxes of each tax regime.

The self-employed are a very interesting group to study *per se*. They have become more numerous and important in recent years, through the rise of platforms such as Uber, Air BnB, or Task Rabbit, and the outsourcing of jobs previously done in-house. In recent work, Katz and Krueger (2016) and Katz and Krueger (2017) cast light on the rise of alternative work arrangements – those differing from conventional self-employment and regular employment – and on the ensuing fragmentation of the labor market. This chapter's focus on the self-employed is mainly driven by the following considerations. First, they are typically shown to be much less constrained than wage earners and can more easily adjust their incomes to tax incentives (Saez, 2010; Kleven and Waseem, 2013). This is important if we want to measure how people respond to simpler or more complex tax policies. Second, since the self-employed are their own decision makers, there is a more direct map between their own understanding of the tax system and their response to it. This link is weakened for wage earners, since it may be their company determining their pay structure and responses to taxes, based on its own (presumably, better) knowledge of the tax system. Third, self-employment in France serves as a particularly well-suited quasi-laboratory for studying the effects of tax simplicity and complexity. Indeed, it has a very unique variety of fiscal "regimes"– or modes of taxation of self-employment– which differ not only in their monetary incentives, but also in their degree of tax simplicity. These fiscal regimes have changed significantly over time, offering the opportunity to study learning and dynamic adjustments. They also impact different groups of agents heterogeneously, thus providing valuable policy variation that helps our estimation.

The "standard regime" treats an individual's net business income (revenues minus costs) as taxable income, which is advantageous for businesses with employees, significant investments, or high operating costs. It does, however, come with the most involved and costly tax accounting requirements, which also limit the scope for misreporting. The "simplified regime" cuts down on tax hassle and allows agents to claim a flat-rate rebate as a fraction of revenues instead of reporting their true business costs, which can be very advantageous for agents with low operating costs. The "super simplified" regime further increases tax simplicity by replacing all income taxes and social insurance contributions by a unique – and relatively low – flat rate payment proportional to gross revenues. The simplified and super simplified regimes require that revenues remain below an eligibility threshold. This threshold depends on the type of business activity, and has changed over time. Thus, broadly speaking, the simplified and super simplified regimes are well suited to agents with small and slow-growing activities, with relatively low operating costs and investments, and with strong preferences for tax simplicity.

To study the behavior of these self-employed individuals, Chapter 2 uses individual tax returns data from the French internal revenue service over the period 1994-2015. The tax returns data are combined with additional administrative and large-scale survey data, to yield information on employment, demographics, education and government benefits received. This highly valuable combination of administrative tax data and census-style survey data allows us to study the characteristics of agents who respond differently to tax incentives.

This chapter's first contribution is to provide summary statistics on the self-employed for the period 1994-2015.[4] And to highlight very significant behavioral responses (in terms of regime choice and

[4]The self-employed are on average older than wage earners, more likely to be retired, more educated, more likely to be in high skill occupations, and have higher labor, capital, and total income. They are less likely to receive unemployment or social insurance benefits. The fraction of agents with self-employed income remained stable at around 5% of all tax filers aged 18-65 until 2009 and has risen since then. The fraction of agents who earn only self-employed income remained at 4% until 2009 and has increased sharply since then.

income) to the notches – discontinuities in the monetary payoffs and in tax simplicity – created by the eligibility thresholds. This bunching at the notch is heterogeneous: it is more pronounced for people in higher tax brackets (who stand to gain more from fiscal optimization), people with at least a high school degree, and people with other sizable sources of income, such as salaried income or pension income.

Chapter 2's second contribution is to uncover evidence that the observed bunching in both simple regimes at least partly stems from tax evasion. Three patterns in the data support this claim. First, revenue statements are more often round numbers close to the thresholds than far from them, an indication that the reported figure is more likely to have been forged. Second, in households with two self-employed individuals, the highest earner appears to shift some of her income to her partner as she approaches the threshold. Third, bunching increases over time, and agents react rapidly to a large and salient institutional change whereas they respond slowly to a smaller or more complex change.

The third contribution is to estimate a structural model by fitting data moments to disentangle between pure monetary motives and the quest for simplicity in explaining individuals' attempt to file under the simpler regimes. This estimation confirms the role of tax evasion. But it also shows that tax simplicity is at work in explaining the observed bunching. Moreover, the simplicity motive appears to be particularly important in the super simplified regime, where it is valued at around 500 € (versus less than 100 € in the simplified regime).

This chapter relates to several strands of literature. First the literature on the response of revenues to taxation: Saez (2010) or Chetty, Friedman and Saez (2013) study the responses to the Earned Income Tax Credit in the United States. We contribute to the literature on taxable income elasticities (Gruber and Saez, 2002; Saez, Slemrod and Giertz, 2012), by bringing hassle costs and the quest for simplicity into the picture, and also by exploiting the heterogeneity across self-employed individuals to provide evidence for tax evasion. Second this chapter relates to the many empirical studies of misreporting in response to taxation, especially recent examples of which are Carrillo et al. (2017), Feldman and Slemrod (2007), Pomeranz (2015), and Gordon and Slemrod (2000). We contribute to this literature by exploiting the multiplicity of self-employment regimes, their staggered introduction, and the heterogeneity between individuals (including between individuals that claim the same income level) to identify the motives for misreporting. Third our analysis of how members of the same household jointly optimize (and misreport) their self-employed earnings echoes the analysis of the joint income decisions among wage earners in Eissa and Hoynes (2004), Eissa and Hoynes (2006), and Gelber (2014). Finally this chapter builds on the literature analysing how tax payers respond to costly information with inattention as in Hoopes, Slemrod and Reck (2017) or with behavioral biases as in Lockwood and Taubinsky (2016) and Lockwood (2016). To infer evidence of tax evasion, we exploit how individuals' learning reacts differently over time to differences changes in self-employment regimes.

Chapter 3 takes a macro perspective to analyse the economic environment in which firms operate. The economic conditions, and in particular the financial conditions, that the firms face, are influenced by the cyclical state of the economy and have an impact on their decisions.

The cyclical state of an economy is traditionally defined as the difference between output and the maximum level of economic activity that can be durably sustained (potential output), this difference being the output gap. Since Okun (1962) documented the empirical link between the evolutions of unemployment and output, many models aiming at estimating the output gap use a few macro variables that have been widely shown to provide information on the cyclical position of an economy: inflation, unemployment, or a survey question relating to the capacity utilisation rate. However, historical evidence suggests that unsustainable developments in the financial and housing sectors can generate large imbalances in the economy even if inflation and unemployment are low and stable (see e.g. White (2006); Hume and Sentance (2009); Schularick and Taylor (2012); Jordà et al. (2013)). The literature began incorporating financial variables into output gaps models to cast light on the Great Recession of 2008-2009, so does not have enough hindsight to understand all the issues in depth. Chapter 3 thus adds financial variables to three models – 2 Unobserved Components Models (UCM) and a semi-structural vector autoregression model (SSVAR) – frequently used in the literature. And it analyses what these variables bring to the output gap estimates: what is the contribution of the financial shocks to the estimated output gaps, the sensitivity of the results to pre-treatment methods, or the real time performance of the models. Finally, the overall accuracy of the estimated output gaps are assessed using receiver operating characteristic analysis based on their capabilities in predicting recessions. The estimations are carried out on eight advanced economies: Canada (CA), France (FR), Germany (DE), Italy (IT), Japan (JP), Spain (ES), the UK, and the US.

The first UCM follows a novel "reduced form" approach pioneered by Borio et al. (2017) to estimate the "finance-neutral" output gap. In this "extended Hodrick-Prescott (HP)" model, pre-transformed financial cycle indicators are directly incorporated as additional covariates into the state-space representation of the univariate HP filter. The second UCM is a simple semi-structural model featuring a Phillips curve, an Okun's law, a stochastic process relating output gap to capacity utilisation, and a separate block that relates financial cycles to the output gap. Both the reduced form and the semi-structural UCMs augmented with financial variables are gaining increasing popularity (for instance in Melolinna and Tóth (2019), upon which our semi-structural UCM is based), however, their relative advantages and limitations have not yet been empirically assessed.

The third model is a "semi-structural vector autoregression (VAR)" model with long-run restrictions. The model is a modified version of Blanchard and Quah (1989) in which we exploit a wider set of information by using several (business and financial) cycle indicators without imposing further restrictions. Since the potential growth builds upon supply shocks only, one set of constraints, stating that only supply shocks have permanent effects on the level of GDP in the long-run, is enough to recover the trend. We formally show that further restrictions are not needed as the other structural shocks are not interpreted and do not need to be separately identified. Furthermore, we show that the limited set of constraints used is the only one relevant for the decomposition of the output gap into contributions of the observables (see Andrle (2013) for a general explanation of the decomposition technique). Very close to our approach, Benati (2012) estimates finance-neutral output gaps using a semi-structural VAR with unemployment (for the US) or the consumption/output ratio (for the UK, Europe and Japan), along with the spread between the 10-year bond yield and the 3-month Treasury bill rate, and the log-difference of real M1. The literature using a structural (and not semi-structural) VAR to estimate (finance-neutral) output gaps relates closely to our analysis, as we show that, to the extent that potential output is built with the shocks having a long-term impact on output, the other restrictions used to identify the other shocks have no impact on potential output and the output gap[5].

Since the semi-structural VAR involves entirely different mechanisms to decompose the trend from the cycle than the two other UCMs, it can provide particularly informative additional insights compared to the other models. All the more so as the ability of the usual univariate or multivariate statistical filters to accurately differentiate between supply and demand shocks has recently been challenged. Coibion et al. (2018) argue that models relying on smoothing techniques gradually but persistently respond to all kinds of shocks to the GDP. The authors show that the Blanchard and Quah (1989) approach, which explicitly distinguishes between temporary and permanent shocks, does not suffer from this shortcoming, and can generate real-time estimates of potential output that are consistent with theoretical predictions much more successfully.

The overall picture underlines the importance of taking financial variables into account when assessing the cyclical position of the economy. Independently of the model considered, the model augmented with financial variables proves to be consistently more effective in identifying unsustainable economic growth paths and in predicting recessions both in-sample and out-of-sample. In ES, the UK and (to a much lesser extent) the US, the credit and house prices boom in the run-up to the Great Recession is clearly identified, as well as a previous boom-bust in the UK during the 1980s-90s. In JP both credit and house prices booms at the turn of the 1990s led to the well-studied and prolonged crisis (the "Lost Decade"). In FR the models signal a house prices boom during the 2000s, but there is no sign of credit boom during the estimation period. Finally, the financial cycles appear relatively muted in CA, DE and IT. At the same time, our results suggest that financial information generally worsens the real-time stability of the models in both the short-run and the longer-run, sometimes even for countries without clear financial cycles identified ex-post (most notably for DE). In other words, the models with financial variables are generally more successful in signalling booms in real-time, but future revisions of the point estimates are larger.

Turning to the general conclusions one can draw for these three models, the approach proposed byBorio et al. (2017) is particularly appealing for its simplicity and real-time stability, but it is sensitive

[5] Said differently, even if these other restrictions do not convincingly disentangle the other shocks, one can still be confident that the output gaps and the potential outputs from these models only obey the long-run restriction à la Blanchard-Quah. See for instance Chen and Gornicka (2020) and Furlanettoy et al. (2020).

to the pre-treatment of the indicators and predicts less reliably future recessions than the two other models. The semi-structural UCM produces, in general, more accurate finance-neutral output gaps with better early warning capabilities than the extended HP filter. The semi-structural VAR provides valuable insights thanks to its distinct trend-cycle decomposition technique, and it captures generally more successfully macroeconomic and financial imbalances than the UCMs, with the best forecasting performance of recession probabilities. It is however sensitive to the trend removal technique applied to the cycle indicators and may react erratically to new observations.

Chapter 2

The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports

The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports

PHILIPPE AGHION ANTONIN BERGEAUD

MATTHIEU LEQUIEN MARC J. MELITZ

Addresses - Aghion: Collège de France, LSE and PSE; Bergeaud: Banque de France and CEP; Lequien: Insee, Banque de France and PSE; Melitz: Harvard University and NBER. A previous version of this paper was circulated under the title: "The Impact of Export on Innovation: Theory and Evidence". The authors would like to thank, without implicating Daron Acemoglu, Ufuk Akcigit, Pol Antras, Kirill Borusyak, Elhanan Helpman, Giammario Impullitti, Thierry Mayer, Isabelle Mejean, Torsten Persson, Stephen Redding, Thomas Sampson, Dan Trefler, John Van Reenen, Daniel Xu, and participants at the 2017 AEA meeting, the NBER Summer Institute 2017, CIFAR and numerous seminars for useful discussions and advises. We are grateful to Connie Lee for excellent research assistance. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-EQPX-17 - Centre d'accès securisé aux données – CASD)

Abstract

We analyze how demand conditions faced by a firm in its export markets impact its innovation decisions. To disentangle the direction of causality between export demand and innovation, we construct a firm-level export demand shock which responds to aggregate conditions in a firm's export destinations but is exogenous to firm-level decisions. Using exhaustive data covering the French manufacturing sector, we show that French firms respond to exogenous growth shocks in their export destinations by patenting more; and that this response is entirely driven by the subset of initially more productive firms. The patent response arises 2 to 5 years after a demand shock, highlighting the time required to innovate. In contrast, the demand shock raises contemporaneous sales and employment for all firms, without any notable differences between high and low productivity firms. We show that this finding of a skewed innovation response to common demand shocks arises naturally from a model of endogenous innovation and competition with firm heterogeneity. The market size increase drives all firms to innovate more by increasing the innovation rents; yet by inducing more entry and thus more competition, it also discourages innovation by low productivity firms.

JEL codes: D21, F13, F14, F41, O30, O47

Keywords: Innovation, export, demand shocks, patents.

2.1 Introduction

Among its many impacts, the Covid crisis dramatically shrinks international trade flows. Other events or policies such as trade wars, though clearly not as extreme, also impact trade. Beyond their immediate impacts, these trade shocks can have long-run consequences, in particular on firms' innovation, one of the main driver of long-run economic growth.[1] The economic magnitude of this link is substantial. In our more conservative specification, we find that a 1 percent expansion/contraction in export demand leads to 52 additional/fewer priority patents (corresponding to the first patent publication for an invention) in the French manufacturing sector – a .64 aggregate elasticity. We analyze how the quantity and quality of this innovation response unfold over time and varies across firms with different initial levels of productivity.

In order to analyze those patenting responses, we merge comprehensive patent records with exhaustive firm-level production and customs data, which cover the whole population of French manufacturing firms. The combined use of these datasets has been made possible by a new algorithm developed in Lequien et al. (2019) that matches a French firm's name with its unique identifier (*Siren*) used in all French administrative business records and allows us to link the innovation activities of a firm with the other firm data sources.

We measure innovation by the flow of *priority* patent applications. All subsequent filings of the same intellectual property (in particular if they are filed at patent authorities in other countries) are secondary filings. We focus on priority patents for two reasons. First because our goal is not to measure a response in patenting but a response in innovation. By focusing on priority patents, we concentrate on patents that correspond to new inventions. Second because we want to avoid capturing the fact that firms that are more involved in international trade are more likely to patent many secondary filings so as to protect their invention in their sales' destinations.

Our first finding is that on average firms respond to a positive export demand shock by innovating more. In other words, we find a significant *market size effect* of export demand shocks on French firms' innovation. Since our specifications always control for sector-year effects, this innovation response must be driven by differences in firm-level innovation responses to demand shocks within each sector. This stands in sharp contrast to the literature measuring sector-wide innovation responses – whether across sectors or for a given sector over time.[2]

Our second finding is that the innovation response to a positive export demand shock takes 2 to 5 years to materialize. In contrast, we find that the response of sales and employment is immediate. We

[1]For a survey on the short-run costs of the 2018 trade war, see Amiti et al. (2019).

[2]In an influential study, Acemoglu and Linn (2004) measure the sector-wide innovation response of the pharmaceutical industry to changes in demand over time.

interpret this difference as a confirmation that the response to export demand shocks captures a market size effect.

Our third finding is that the impact of a positive export demand shock on innovation is entirely driven by French firms with above median productivity levels (in an initial period prior to the demand shocks). This heterogeneous response could simply reflect the fact that the demand shock only affects the most productive firms. We check that this is not the case by allowing for a different impact of the export demand shocks on sales or employment depending upon initial productivity levels. We find that in contrast to what we observe for innovation, there is no heterogeneous response of sales or employment to a demand shock for low versus high productivity firms. Thus, similar demand shocks only lead to future innovation responses by relatively more productive firms.

These results provide some additional context to the recent literature documenting the rise of superstar firms: the skewed innovation response is likely to generate further increases in market share for the best performing firms leading to increases in market concentration. Indeed, Autor et al. (2017) document that this growth in concentration is most apparent in industries with above average growth in patent-intensity.

Our identification strategy relies on the construction of firm-level demand shocks that respond to aggregate conditions in a firm's export destinations but are independent of firm-level decisions (including the concurrent decisions for export-market participation and the forward looking innovation response). Following Hummels et al. (2014), this type of export demand shock has been used extensively in the recent empirical trade literature. It leverages detailed information on the set of products exported to specific destinations by a firm at a prior given date (prior to any changes in innovation that we analyze in our sample). Focusing on this export-driven measure of market size means that we are abstracting from the potential effects of domestic market demand on firms' innovation. For this market, we cannot separate out the causal effects of domestic market size on innovation from the reverse effect of innovation on domestic demand and market size.

We show that our results using this identification strategy are robust to many different specifications including variations in the measure of and functional forms for innovation. We also perform placebo tests that independently confirm that our causation inference from increases in market size to innovation are well founded.

While several explanations might be entertained to explain why the effect of export on innovation should be skewed towards more frontier firms, we show that this outcome arises naturally from a model of exports and innovation with endogenous innovation and markups. In this setting, a positive export demand shock induces not only a direct market size effect – which increases innovation for all firms – but also a competition effect. The idea is that an increase in market size in any export destination will attract new firms into the export market as more firms find it profitable to sell there. And indeed we find a positive correlation between our export demand shocks and various measures of firm entry into the corresponding destination markets. With endogenous markups (linked to endogenous price elasticities), this competition effect associated with entry impinges disproportionately on the market share of the less productive firms, reducing their incentives to innovate. Overall, this combination of the direct market size effect and of its induced competition effect leads to a skewed innovation response between more and less productive firms. Firms closest to the technological frontier increase innovation the most, while the combined effect can even be negative for the least productive firms.

Our analysis relates to several strands of literature. There is first the theoretical literature on trade, innovation and growth (see Grossman and Helpman, 1991a,b, Aghion and Howitt, 2009, chapter 13, and more recently Akcigit et al., 2018).[3] Our paper also relates to the recent empirical literature on firm-level trade and innovation. In particular both Lileeva and Trefler (2010) and Bustos (2011) highlight a clear relationship between R&D efforts and export status. Our analysis contributes to this literature in two main respects: (i) this literature focuses on the *extensive* margin of export markets (i.e. whether a firm exports or not to a particular market or set of markets) whereas we consider instead the effect of the *intensive* margin of exports (i.e. of the size of export markets) on innovation;[4] (ii) we use innovation outcomes - the flow of priority patent filings - instead of R&D spending as our main measure of innovation, whereas these papers consider the causal impact of new export markets on R&D spending.[5]

There is also a recent literature on trade and innovation that focuses on the impact of import competition on domestic firms (see Bloom et al., 2016; Iacovone et al., 2011; Autor et al., 2016; Bombardini et al., 2017). These papers investigate whether import competition induces firms to innovate more in order to escape competition as in Aghion et al. (2005). Empirically, our work is quite distinct as we examine the market expansion channel related to exports. Our theoretical model therefore does not feature an escape competition channel: reductions in market share generate reductions in innovations, though disproportionately so for low productivity firms.

Finally, our work contributes to the empirical literature on market size and innovation, starting with Acemoglu and Linn (2004). We add to this literature in three main respects: (i) by providing evidence of a widespread (manufacturing) firm-level market size effect that is not driven by any sector-level dynamics; (ii) by showing that this market size effect is skewed and mainly driven by the most productive French firms; (iii) by looking at the time dynamics of the market size effect of expanded

[3] Akcigit et al. (2018) develop and calibrate a new dynamic trade model where firms from different countries compete strategically for leadership in domestic import and export markets. Their model predicts that trade openness encourages innovation in advanced sectors and discourages it in backward sectors. Dhingra (2013) and Impullitti and Licandro (2018) also develop theoretical models with endogenous firm innovation and endogenous competition (via endogenous markups). Dhingra (2013) focuses on the firm-level trade-offs between innovation and product variety, whereas Impullitti and Licandro (2018) focuses on the consequences of innovation for growth and welfare.

[4] Restricting attention to the extensive margin makes it somewhat more difficult to analyze the details of how the market size channel operates: one reason being that several aspects are changing for a firm as it makes the big step of becoming an exporter.

[5] In related work, Coelli et al. (2016) document the patenting response of firms in response to the Uruguay round of tariff levels.

export markets on firm-level innovation: in particular we show that while a positive export demand shock immediately increases the firm's sales, the innovation response takes several additional years to materialize in new patents. However, one should keep in mind that our analysis is grounded in the market size variations arising from export destinations, which means that we do not use variations coming from the domestic markets. Thus we leave open the question as to whether the domestic market size affects firms in a similar way as the export market size effect that we document.

The remaining part of the paper is organized as follows. Section 2.2 presents the data and shows some descriptive statistics on export and innovation. Section 2.3 describes our estimation methodology. Sections 2.4, 2.5 and 2.6 present our empirical results respectively regarding the effect of market size on innovation, its heterogeneous impact with productivity and falsification tests. Section 2.7 develops a model of export and innovation featuring both a direct market size and an induced competition effect, which predicts that the innovation response to a positive export shock is skewed towards the more productive firms. Section 2.8 concludes.

2.2 Exporters and innovators: data and descriptive statistics

In this section, we briefly present our datasets and show some descriptive evidence. Further details about data construction can be found in Appendix 2.A.

2.2.1 Data sources

Our goal is to explore information on French firms' exports to capture variations in their market size that we can connect to innovation (patenting) outcomes. We also want to look at how this relationship varies across firms with different levels of productivity. Toward this goal, we build a database covering all French firms by linking export, production and innovation data from 1994 to 2012. Our database draws from three sources: (i) French customs, which reports yearly export flows at a very disaggregated product level (representing over 10,000 manufacturing products) by destination; (ii) administrative fiscal datasets (`FICUS` and `FARE` from Insee-DGFiP), which provide extensive production and financial information for all firms operating in France; (iii) the Spring 2016 vintage of `PATSTAT` patent dataset from the European Patent Office, which contains detailed information on all patent applications from many patent offices in the world. In our analysis we will focus on patent applications by French firms, regardless of the origin of the patent office (see below and Appendix 2.A for details).

Matching patents and firms: Although each French firm has a unique identifying number (*Siren*) across all French databases, patent offices do not identify firms applying for patents using this number but instead use the firm's name. This name may sometime carry inconsistencies from one patent to another and/or can contain typos. Various algorithms have been developed to harmonize assignees' names (see Morrison et al., 2017 for a review) but none of those have been applied specifically to French firms. One notable exception is the rigorous matching algorithm developed in Lequien et al. (2019) to link each patent application with the corresponding French firms' Siren numbers, for all firms with more than 10 employees. This new method, based on supervised learning and described in Appendix 2.A.4, provides significant performance improvements relative to previous methods used in the empirical patent literature: based on a verification sample similar to the learning sample, its recall rate (share of all the true matchings that are accurate) is 86.1% and its precision rate (share of the identified matches that are accurate) is 97.0%. This is the matching procedure we use for our empirical analysis in this paper.

Measure of innovation: Our main measure of innovation consists of a count of *priority* patent applications. This corresponds to the first patent publication that describes an invention. All subsequent filings of the same intellectual property in other jurisdictions (for example in order to extend the geographical coverage of the protection) are secondary filings. We make this restriction for two reasons. First because our goal is not to measure a response in patenting but a response in innovation. By focusing on priority patents, we concentrate directly on patents that correspond to new inventions. Second because we want to avoid capturing the fact that firms that are more involved in international trade are more likely to patent many secondary filings so as to protect their invention in the markets they export to. Priority patents correspond to 35% of the total set of patents but 95% of innovative firms (firms that hold any patent, whether a priority or a secondary filing) in our sample hold at least one priority patent. This suggests that most of the patents we observe in the data are successive secondary filings of the same innovation by the same firm, and legitimate the use of priority applications as our main measure of innovation. Appendix 2.A provides additional details on the construction of our patent measures. For robustness, we report all of our main results using an alternative patent measure based on citation weights for all patent applications by a firm (citations received within a 5 year window). Following Hall et al. (2005), this measure has been widely used in the literature to more accurately capture the innovative relevance of patents. We have also confirmed that our results are robust to a much wider set of patent measures in Appendix 2.C (see in particular Figures 2.15).

Capturing variations in market size: Finally, to capture variations in firms' market size, we use CEPII's `BACI` database of bilateral trade flows at the HS6 product level (covering more than 5,000 manufacturing products, see Gaulier and Zignago, 2010) to construct measures of demand shocks across export destinations. These data cover the period 1995-2012.

Sample restrictions: Although our main firm-level administrative data source is comprehensive, with more than 46.8 million observations spanning nearly 7.5 million different firms from 1995 to 2012, we restrict our data sample for several reasons. First, we restrict our attention to private business corporations (legal category 5 in the INSEE classification). We thus drop state-owned firms, self-employed businesses, and non-profit organizations as we focus on profit-maximizing firms. Second, we drop firms with less than 10 employees since our matching to the patent data is substantially less complete for those firms (as we previously described). These two restrictions substantially reduce the number of firms in our sample. Yet, the bulk of aggregate employment (77%), sales (80%), and exports (92%) remain in our sample. Those firms are matched with 460,000 patents in PATSTAT, including 170,000 priority patents. Lastly, since our detailed customs trade data only covers goods trade (and not services), we will further restrict our sample to the manufacturing sector.[6] This reduces our working sample to 66,679 firms. Nevertheless French aggregate exports and innovation are still concentrated in manufacturing covering 55% of aggregate exports and 43% of patents. Table 2.1 summarizes these successive sample restrictions and also shows the average number of firms operating in any given year of our sample. For our manufacturing sample, we see that this represents 42,924 firms on average per year between 1995-2012.

Table 2.1: Successive restriction of the sample

	Total Firms	Firms per Year	Employment	Sales	Exports	Patents
Full	7,474,147	2,597,852	100	100	100	
Private business Corp.	2,888,647	1,114,651	88	90	97	
More than 10 emp	400,662	260,386	77	80	92	100
Manufacturing	66,679	42,924	19	20	55	43

Notes: This Table gives the number of distinct firms and average number of firms per year as well as the share of employment, sales, exports and patents in each sample as compared to the Full (raw) firm level dataset (in %). All columns except the first consider yearly average over the period 1995-2012. Full correspond to our complete sample of firms based on administrative data (see Section 2.2). "Private Business Corp." corresponds to this sample restricted to firms that are in Legal category ("*catégorie juridique*") number 5. "More than 10 emp" further reduces the sample to firms that are at least once over 10 employees over the period of observation. "Manufacturing" restricts to firms that are always classified in a manufacturing sector.

The case of multinational groups: Our dataset does not allow us to properly take into account the case of multinational groups, an issue which often arises when dealing with national firm level data. The presence of multinational groups tends to break the relationship between export shocks and patenting since these groups may locate their R&D activities in different countries than the location of production. In particular, the R&D activity for production based in France may be located elsewhere under a different entity of a multinational's group. Conversely, the R&D activity can be located in France whereas a substantial share of production is outsourced in other countries. In these cases, we will not record the appropriate link between the export shocks for this producer and an induced innovation (patents). This measurement issue works against our obtained results of a positive response of patenting to export shocks that is increasing with a firm's proximity to its industry frontier. Thus,

[6] Although the customs data also covers the wholesale sector, we also exclude those firms as they do not produce the goods that they export.

we conjecture that our results would be strengthened if we had the needed information to exclude broken production/R&D links amongst the multinational groups in our sample.

2.2.2 Sector breakdown and skewness

Starting from our sample of manufacturing firms from Table 2.1, Table 2.2 shows how those firms are distributed across sectors (using the NACE 2-digit classification), along with their average employment and sales per firm over our sample period from 1995-2012 – shown as yearly averages.[7] Table 2.2 also shows the proportion of exporters and innovators (firms with at least one patent) in each sector (again, averaged over our sample years) – along with the average exports per exporter (firms with positive exports) and the average number of patents and priority patents per innovator. We clearly see that innovators represent a small minority of manufacturing firms. Only 2.7% of firms introduce any new patents in any given year (on average). Looking across years, 9.7% of firms have at least one patent in one of those years. This is the set of firms we will classify as innovators in our ensuing analysis. Although a minority of firms, they nevertheless represent 37% of employment, 45% of sales, and 60% of exports for the manufacturing sector. In Table 2.3, we report the same statistics for employment, sales, exports, and patents as sector-level shares. We see that priority patents are concentrated in the computer and electronic, machinery and equipment, and motor vehicles sectors, jointly accounting for 44.4% of the priority patents in manufacturing.

Table 2.2 reveals that the number of patents introduced each year by innovators can be substantial – especially in some sectors. There is a huge amount of dispersion underlying that average number of patents. To highlight this skewness, we show the Lorenz curve for the distribution of those patents in Figure 2.1, along with the Lorenz curves for exports, sales, and employment in one of our sample years (2007). Figure 2.1 confirms the previously reported finding that firm-level exports are significantly more skewed than sales and employment (e.g. see Mayer and Ottaviano, 2008 and Bernard et al., 2018): 1% of firms account for 70% of aggregate exports in 2007, whereas the top 1% of firms based on total size account for 51% of sales (ranked by sales) and 33% of employment (ranked by employment). But Figure 2.1 also shows that patenting is even significantly more skewed than exporting: 1% of all firms account for 91% of priority patents in 2007. (Although we don't show the Lorenz curve for citations, it is even more skewed than that for patenting: all the 5-year citations are owned by the top 1.6% of firms). Yet, these univariate statistics for patenting and exporting do not capture the massive overlap between these two activities across firms – which we investigate in more detail below.

[7] Throughout, we define sectors at the 2-digit level of the European NACE rev2 classification. We also eliminate the tobacco sector (# 12) as it only contains two firms.

Table 2.2: Exports and innovation in the manufacturing sector

			Mean per Firm		%	Mean per Exporter	%	Mean per Innovator	
Sector	Description	Firms	Employment	Sales	Exporter	Exports	Innov.	Patents	Priority
10	Food products	6,612	49	13	25	7	0.4	5.4	2.2
11	Beverages	397	70	34	71	14	*	*	*
13	Textiles	1,613	42	6	63	3	2.3	3.5	1.8
14	Wearing apparel	1,579	39	5	54	3	0.5	1.8	1.4
15	Leather	491	60	8	60	5	1.0	1.7	1.1
16	Wood	1,922	30	4	41	2	0.7	1.5	1.2
17	Paper	2,385	52	11	49	5	1.6	4.7	2.0
18	Printing	1,361	26	4	26	1	0.5	3.1	1.9
19	Coke	112	338	738	71	168	7.8	57.9	17.3
20	Chemicals	978	106	37	80	18	6.0	9.9	4.1
21	Basic pharmaceutical	298	224	91	79	42	11.8	15.5	3.0
22	Rubber and plastic	2,367	78	13	64	5	5.0	5.2	2.8
23	Other non-metallic	1,615	67	14	42	5	2.7	11.3	3.3
24	Basic metals	1,125	91	24	54	17	3.0	5.2	2.1
25	Fabricated metal	7,655	34	5	39	2	1.7	3.2	1.9
26	Computer and electronic	2,318	89	18	59	11	7.6	9.0	4.6
27	Electrical equipment	527	156	33	69	17	8.6	18.0	9.2
28	Machinery and equipment	3,263	93	27	63	10	7.4	5.7	3.2
29	Motor vehicles	941	126	40	55	27	4.1	22.8	21.9
30	Other transport equipment	422	192	58	59	42	7.7	20.2	10.1
31	Furniture	985	38	5	41	1	1.1	1.9	1.5
32	Other manufacturing	1,008	47	8	54	7	4.0	19.1	9.5
33	Repair of machinery	2,952	27	3	23	1	1.0	4.0	1.9
	All Manufacturing	42,924	58	14	45	8	2.7	8.0	4.1

Notes: This table presents the number of firms, average employment, sales, employment and exports (sales and exports are in million of Euros, employment in number of employees), the share of exporters, the total number of patents and of priority patents in the sector and the share of innovators. Sector codes correspond to the 2 digit NACE classification. The data presented represents the yearly averages from 1995 to 2012. Cells with too few observations to ensure data confidentiality are replaced with $*$.

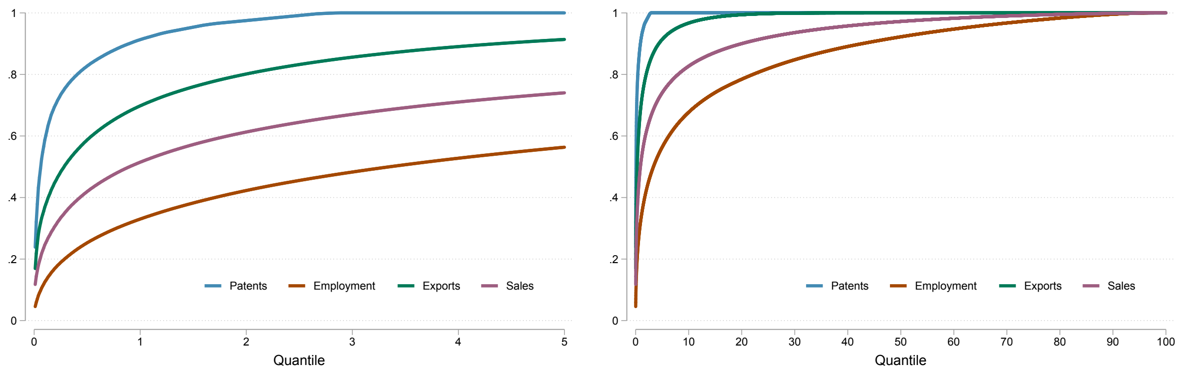
Table 2.3: RELATIVE IMPORTANCE OF EACH SECTOR

NAF	Description	Share of total (in %)					
		Firms	Employment	Sales	Exports	Patents	Priority
10	Food products	15.5	13.0	14.3	7.7	1.5	1.1
11	Beverages	0.9	1.1	2.2	2.5	*	*
13	Textiles	3.7	2.7	1.6	2.3	1.4	1.4
14	Wearing apparel	3.6	2.4	1.2	1.4	0.1	0.2
15	Leather	1.1	1.2	0.6	0.8	0.1	0.1
16	Wood	4.5	2.3	1.4	0.8	0.2	0.3
17	Paper	5.4	4.4	3.6	3.2	1.5	1.4
18	Printing	3.3	1.5	0.8	0.2	0.2	0.3
19	Coke	0.3	1.3	8.7	5.5	4.0	2.8
20	Chemicals	2.3	4.2	6.0	8.8	6.6	4.7
21	Basic pharmaceutical	0.7	2.6	4.2	5.7	5.6	1.9
22	Rubber and plastic	5.5	7.4	5.1	4.9	6.6	7.1
23	Other non-metallic	3.8	4.3	3.6	2.3	5.1	3.1
24	Basic metals	2.6	3.8	3.8	5.9	1.8	1.4
25	Fabricated metal	18.0	10.6	5.9	4.0	4.6	5.4
26	Computer and electronic	5.3	7.7	6.1	9.1	15.2	15.1
27	Electrical equipment	1.3	3.1	2.8	4.1	7.7	8.3
28	Machinery and equipment	7.5	12.1	15.1	13.5	14.5	16.4
29	Motor vehicles	2.2	4.5	5.0	6.6	7.5	12.9
30	Other transport equipment	1.0	3.1	4.0	6.9	7.0	6.8
31	Furniture	2.3	1.5	0.8	0.4	0.2	0.4
32	Other manufacturing	2.4	1.9	1.4	2.6	7.2	7.5
33	Repair of machinery	7.0	3.3	1.6	0.6	1.3	1.2

Notes: This table presents the share of value added, employment, export and patents (all patents and priority patents) accounted for by each 2-digit manufacturing sector as well as the share of firms in each sector. Data are averaged over the period 1995-2012. Cells with too few observations to ensure data confidentiality are replaced with $*$.

Figure 2.1: LORENZ CURVES FOR PRIORITY PATENTS, EXPORTS, SALES AND EMPLOYMENT

(a) Top 5 percentiles (b) Whole distribution



Notes: Lorenz curves plot cumulative distribution function for priority patents, employment, export and sales. Data are for manufacturing firms and for the year 2007.

2.2.3 The nexus between innovation and exports

Looking across our sample years (1995-2012), Table 2.4 reports different size-related performance measures (averages per firm) based on their exporter and innovator classification. As we previously discussed, we classify firms as innovators if they introduced at least one patent during those sample years. From here on out, we classify exporters in a similar way as a firm with positive exports in at least one of our sample years. This raises the proportion of exporting firms to 61% of our manufacturing sample (45% of firms export on average in any given year, c.f. Table 2.2). Table 2.4 confirms the well-documented size differential in favor of exporters. However, several new salient features regarding innovators pop-out from this table. First, innovating firms are massively concentrated among exporters: only 5% of innovators do not report any exporting. Second, non-exporting innovators do not look very different from non-exporting non-innovators, and the various measures of firm size (employment, sales, value-added) respectively for innovators and non-innovators among non-exporters remain close to each other;[8] and third, these same measures of firm size differ markedly between innovators and non-innovators among exporters: innovators employ on average 4.5 times more workers and produce 7-8 times more output and value-added than non-innovating exporters. They export almost 10 times more than non-innovators and reach more than three times the number of export destinations. These size differentials are several times larger than those between exporters and non-exporters. In the aggregate, this small subset of innovators accounts for over half of French manufacturing exports.

[8]This is not the case outside of the manufacturing sector. In those other sectors, non-exporting innovators are substantially bigger than their non-exporting and non-innovating counterparts. We conjecture that this is driven by the fact that exporting no longer serves the same performance screening function outside of manufacturing.

Table 2.4: EXPORTERS AND INNOVATORS ARE BIGGER

	Non-exporter		Exporter		Total
	Non-innovator	Innovator	Non-innovator	Innovator	
Firms	13,266	173	25,045	4,440	42,924
Employment	20	19	51	215	58
Sales	4.2	2.4	10.7	62.3	14.1
Value Added	0.8	0.9	2.7	14.7	3.4
Export	0	0	2.4	20.8	3.6
Countries	0	0	4.8	17.3	4.6
Products	0	0	5.0	16.1	4.6

Notes: This table presents basic descriptive statistics across four categories of manufacturing firms whether they innovate, export, both or none. Employment is given in full-time equivalent on average over the year and exports, sales and value added are in million euros. Countries is the number of destination countries for exports. Employment, Sales, Value Added, Age, Exports, Countries and Patents are taken as a yearly average over the whole period 1995-2012.

Table 2.5: EXPORT AND INNOVATION PREMIA

Panel 1: Premium for being an exporter (among all manufacturing firms)					
	(1)	(2)	(3)	Obs.	Firms
log Employment	0.865	0.843		754,008	66,563
log Sales	1.361	1.344	0.463	764,372	66,601
log Wage	0.122	0.100	0.113	752,774	66,548
log Value Added per Worker	0.209	0.184	0.183	744,076	66,119
Panel 2: Premium for being an innovator (among all exporting manufacturing firms)					
	(1)	(2)	(3)	Obs.	Firms
log Employment	1.001	0.976		519,463	42,023
log Sales	1.270	1.239	0.205	525,674	42,042
log Wage	0.118	0.096	0.111	518,682	42,019
log Value Added per Worker	0.207	0.183	0.185	512,040	41,795
log Export Sales (Current period exporters)	2.015	1.897	0.790	346,273	41,659
Number of destination countries	12.55	11.47	6.95	530,729	42,082

Notes: This table presents results from an OLS regression of firm characteristics (rows) on a dummy variable for exporting (upper table) or patenting (lower table) from 1994 to 2012. Column 1 uses no additional covariate, column 2 adds a 2-digit sector fixed effect, column 3 adds a control for the log of employment to column 2. Wage corresponds to total payroll divided by employment. All firm characteristic variables are taken in logs. All results are significant at the 1 percent level. Upper table uses all manufacturing firms whereas lower table focuses on exporting manufacturing firms.

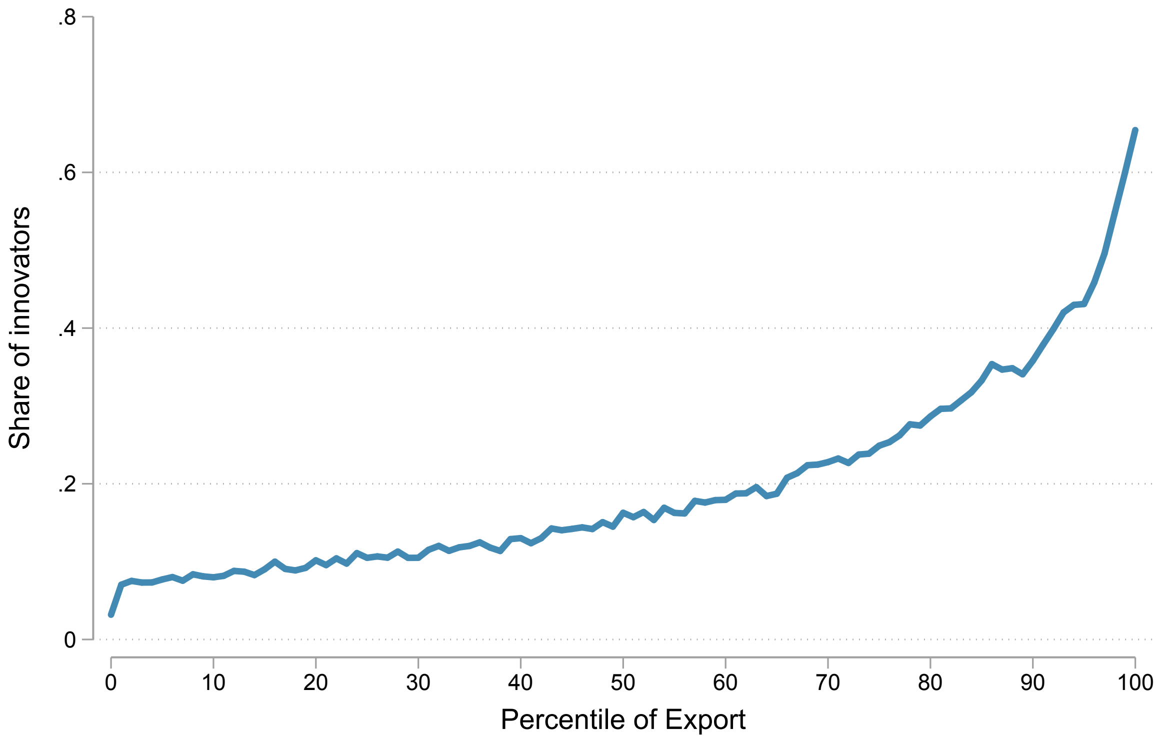
In order to compare exporters to non-exporters and innovators to non-innovators, within specific groups, we compute export and innovation premia (in log points). Consider first the exporter premia reported in the top panel of Table 2.5. These premia are generated by regressing the performance measure of interest (listed in the rows) on our exporter indicator – with each cell representing a separate regression. Column 1 includes no other controls; Column 2 adds a 2-digit sector fixed effect (see Table 2.2); and Column 3 controls for firm employment, in addition to the sector fixed effect. Since we are using a broad cross-year definition for exporter status, we expect these premia to be lower than measures based on current-year exporter status since firms who drop in and out of export markets tend to be substantially smaller than year in year out exporters. This is the case for the premia in column 1 compared to similar numbers reported by Bernard et al. (2018) for U.S. firms in 2007. Yet, once we control for sectors in column 2, the reported premia become much more similar. In particular, we find that even within sectors, exporters are substantially larger than non-exporters. And we also find that large differences in productivity and wages in favor of exporters persist even after further controlling for firm employment.

In the bottom panel, we focus on the subset of exporters from the top panel, and report the *additional* premia in favor of innovators within this subset. As with the top panel, those premia are calculated by running separate regressions on our innovator indicator. Even within this subset of bigger and better performing firms, innovators stand out: they are substantially bigger, more productive, and have larger total wage bill. They also export substantially more (and to more destinations) than non-innovative exporters. All these differences persist within sectors and controlling for firm employment.

Even these large premia do not fully reflect the concentration of innovative and exporting activities within the more restricted subset of firms that are both exporters and innovators. Figure 2.2 plots the share of innovating firms for each percentile of the firm export distribution. We see that the innovative firms are highly concentrated within the top percentiles of the export distribution. At the 80^{th} percentile of the export distribution, 30% of the firms have some patenting experience. And the increase in the share of innovative firms with the percentile of the export distribution is highly convex. Above the 95^{th} percentile of the export distribution, a majority of firms are innovators; in the top percentile, 66% of the firms are innovators. Those firms in the top export percentile account for 41% of the aggregate share of French patents.[9]

[9]Of course part of the relationship in Figure 2.2 could be driven by a scale effects: large firms tend to export more and are more likely to innovate. When we rank firms in percentile of export intensity (instead of absolute export) we still find a near monotonic increase in the share of innovators for export intensity in the 5-95% range. After this threshold, the relationship becomes negative as the last 5 percentiles of export intensity are dominated by unusual small firms that export virtually all of their sales.

Figure 2.2: THE SHARE OF INNOVATORS JUMPS AT THE TOP OF THE EXPORT DISTRIBUTION



Notes: Percentiles of exports are computed each year from 1995 to 2012 separately and then pooled together. For each percentile, we compute the share of innovators. Each percentile contains the same number of firms, except for percentile 0 that contains all the firms with no export. Manufacturing firms only.

2.3 Empirical Framework

2.3.1 Firm level export demand shocks

We have just documented a strong correlation between exports and innovation in the cross-section of French manufacturing firms. However, this correlation does not say much about the direction of causation: from innovation to exports (a major innovation leads to growth in export demand and entry into new export markets), or from exports to innovation. Moreover, other firm-level changes could generate concurrent changes in both innovation and exports (for example, a new management team). Thus, to identify the causal relationship from exports to innovation, we need to identify a source of variation in firm exports that is exogenous to changes within the firm (and in particular to the innovation activity of the firm). We follow Mayer et al. (2016) in building an exogenous firm-level measure of export demand shocks.

To construct these export demand shocks, consider a French exporter f who exports a product s (measured at the HS6 level) to destination j at an initial date t_0. Let $M_{j,s,t}$ denote the aggregate import flow in product s into country j from all countries except France at time $t > t_0$. $M_{j,s,t}$ reflects the size of the (s, j) export market at time t. We then sum over the $M_{j,s,t}$ across destinations j and products s weighted by the relative importance of each market (s, j) in firm f's exports at the initial date t_0. The underlying idea is that subsequent changes in destination j's imports of product s from the world (excluding France) will be a good proxy for the change in export demand faced by this firm. By excluding French exports to this destination, we seek to exclude sources of variation that originate in France and may be correlated with changes for the firm.[10]

We then scale the weighted export demand variable by the firm's initial export intensity (at t_0) so that our demand shock scales proportionately with a firm's total production (as a firm's export intensity goes to zero, so does the impact of any export shock on total production).

Formally, t_0 is the first year with positive exports in both customs (to compute destination market shares) and production data (to compute export intensity).[11] X_{f,j,s,t_0} denotes firm f's export flow to market (j, s) at time t_0. The export demand shock for firm f between t and $t-1$ is then constructed as:

$$\Delta D_{f,t} = \sum_{j,s} w_{f,j,s,t_0} \left(\frac{M_{j,s,t} - M_{j,s,t-1}}{\frac{1}{2}(M_{j,s,t} + M_{j,s,t-1})} \right), \tag{2.1}$$

where the weight $w_{f,j,s,t_0} \equiv (X^*_{f,t_0}/S^*_{f,t_0})(X_{f,j,s,t_0}/X_{f,t_0})$ represents firm f's initial share of sales of product s, at the HS6 level, to destination j and $X_{f,t_0} = \sum_{j,s} X_{f,j,s,t_0}$ represents the firm's total exports at date t_0. The asterisks on firm f's initial export intensity $X^*_{f,t_0}/S^*_{f,t_0}$ indicate that the underlying data for total exports X^*_{f,t_0} and sales S^*_{f,t_0} come from the production data (as opposed to customs data which we use to calculate the destination/product specific market shares).[12]

There are some clear outliers in the distribution of this demand shock $\Delta D_{f,t}$ across firms. They typically involve firms that export a small number of often highly specialized products to small destinations (such as yachts to Seychelles and Maldives). In order to deal with these outliers in a consistent way, we trim our demand shock $\Delta D_{f,t}$ at 2.5% (eliminating those trade shocks below/above the 2.5th and 97.5th percentiles in each year). We report our main results on the response of innovation to this trade shock using trimming thresholds between 0-5% in Appendix 2.C (Figures 2.18).

[10]One potential source of endogeneity may arise in markets where a French firm has a dominant position. We check that our results are robust to dropping firm-destination pairs whenever the firm's market share in the destination exceeds 10%. See Figure 2.17 in Appendix 2.C.

[11]This year is 1994 for about half of the firms and is used as a reference year in most of our analysis.

[12]Total exports reported by customs and in the production data do not always exactly match, though they are highly correlated. One potential source of difference comes from small exports towards other European Union countries which are not reported in customs data (see Appendix 2.A for more details).

Demand shock as a shift share instrument: We note that the time variation in our demand shock $\Delta D_{f,t}$ only stems from the variation in the world export flow $M_{j,s,t}$ and not in the firm-level weights, which are fixed at their value in the initial export period t_0. We expect that a firm's innovation response at time $t > t_0$ will induce changes to its pattern of exports at time t and beyond, including both intensive margin responses (changes in exports for a previously exported product s to a destination j) and extensive margin responses (changes in the set of products s sold across destinations j). By fixing the firm-level weights in the initial period t_0 (including the extensive margin set of products and destinations), we exclude this subsequent endogenous variation in exports from our demand shock. This is quite similar to a standard shift-share or "Bartik" (Bartik, 1991) setting in which aggregate shocks are combined with measures of shock exposure. In our case the sum of exposure weights w_{f,j,s,t_0} across (s, j)'s is different from 1 and varies across firm. We follow Borusyak et al. (2021) who argue that in such "incomplete shift-share" case with panel data, one needs to control for this sum interacted with a time dummy in our regressions.[13]

2.3.2 Estimation strategy

Here we spell out the baseline regression equations of French firm's innovation on the export demand shock variables $\Delta D_{f,t}$. Our identifying assumption is that after controlling for any sector-level variation by year and firm characteristics at and prior to t_0, subsequent variations in the firm-level export demand shock are uncorrelated with firm-specific shocks to innovation.

As we have no presumption regarding the timing of this innovation response to demand shocks, we include a full set of lags and leads for the demand shock $\Delta D_{f,t}$ in our regressions. Our identification strategy nevertheless relies on the fact that our shock is independent of previous innovation decisions and we will check that the response of innovation to future shocks remains insignificant – in other words, no pre-trends.[14] We restrict our analysis to the subset of innovating firms (i.e. firms with at least one patent between 1985 and 2012), and check that entry into innovation subsequent to 1994 does not bias our sample.[15] Out of our sample of 66,679 manufacturing firms (see Table 1), there are

[13] Borusyak et al. (2021) discuss the possibility of purging the shift-share instrument from a specific dimension. In our case, we have experimented using the residual of $M_{j,s,t}$ on different sets of fixed effects (product and time, country and time) in the construction of the demand shock (using the same weights as in the baseline). Our results are qualitatively unchanged. Even with the unresidualized shock, the effective number of shocks, measured by the inverse concentration index of the exposure share, is also much smaller than the number of export destination markets (s, j) (see Borusyak et al., 2021 for more details).

[14] A simpler and more naive approach would have been to estimate a static model in which the dependent variable would be directly regressed on the demand shock lagged appropriately. Although such static model delivers consistent predictions with our result, the coefficients would most likely be biased in the sense that the firm's response is also affected by subsequent and previous shocks.

[15] In Appendix 2.C, Figure 2.19 shows that our main results are essentially unchanged when we further restrict the sample to firms who innovated before 1994. Our sample also includes firms for which we can define a t_0, i.e. firms that exported

4,785 such innovators. Not all of them are active throughout our sample period. On average across those years there are 1,159 innovators in our sample (2.7% of 42,924 manufacturing firms operating in a given year).

Our main estimation strategy is described by:

$$\begin{aligned}\Delta Y_{f,t} &= \left(\sum_{\tau=-k'}^{k} \alpha_\tau \Delta D_{f,t-\tau}\right) + \boldsymbol{\gamma} \cdot \boldsymbol{Z}_{f,t_0} + \tilde{\boldsymbol{\gamma}} \cdot \left(\tilde{\boldsymbol{Z}}_{f,t_0} \times \chi_t\right) + \varepsilon_{f,t} \\ &= \boldsymbol{\alpha} \cdot \Delta_k \boldsymbol{D}_{f,t} + \boldsymbol{\gamma} \cdot \boldsymbol{Z}_{f,t_0} + \tilde{\boldsymbol{\gamma}} \cdot \left(\tilde{\boldsymbol{Z}}_{f,t_0} \times \chi_t\right) + \varepsilon_{f,t},\end{aligned} \tag{2.2}$$

where $\Delta Y_{f,t}$ is firm f's outcome of interest between t and $t-1$; $\boldsymbol{Z}_{f,t_0}$ is a vector of controls for firm f at t_0; and $\tilde{\boldsymbol{Z}}_{f,t_0}$ is a subset of that vector, which is interacted with year interval fixed-effects χ_t. The second equation uses the vector notation $\Delta_k \boldsymbol{D}_{f,t} = [\Delta D_{f,t+k'}, \Delta D_{f,t+k'-1}, ..., \Delta D_{f,t}, ..., \Delta D_{f,t-k}]$ and $\boldsymbol{\alpha} = [\alpha_{-k'}, ..., \alpha_k]$. As we previously discussed, we include a sector indicator and the firm's prior export intensity (at t_0) in the subset $\tilde{\boldsymbol{Z}}_{f,t_0}$ of $\boldsymbol{Z}_{f,t_0}$, so those are also interacted with the year dummies.

Our specification in first-difference eliminates any bias that would be generated by a correlation between non time-varying firm characteristics (likely to affect current and future innovation) and the *level* of the demand shock shock $D_{f,t}$.[16] We additionally want to control for a potential correlation between those firm characteristics and future *changes* in the demand shock $\Delta D_{f,t}$. Following Blundell et al. (1999) and Blundell et al. (2002), we use a control function approach based on firm performance variables measured at t_0. We use the levels and growth rates of sales and employment as controls, which we include in the vector $\boldsymbol{Z}_{f,t_0}$. In addition, we include controls for the firm's past and current rate of innovation at t_0 whenever we use an innovation measure as the dependent outcome. We describe the functional form for those additional controls in more detail in the following section. We note that this type of correlation between changes in the demand shock $\Delta D_{f,t}$ and firm characteristics is substantially less likely than a correlation with the level of the demand shock $D_{f,t}$. We have checked that there is indeed a strong correlation between that demand shock in levels and the firm characteristics in our control function (better performing firms tend to export to destinations with higher levels of demand). However, there is no correlation between those variables and changes in demand $\Delta D_{f,t}$.

Lastly, Borusyak et al. (2021) and Goldsmith-Pinkham et al. (2018) point out that even when such a correlation between firm characteristics and future demand shocks remains, the induced bias disappears as the number of shocks (our combination of destination-product pairs) grows large.

at least once since 1994. t_0 is used as a reference year and can be any year from 1994. Figure 2.20 shows that our results hold if we restrict to firms for which $t_0 = 1994$.

[16] As discussed in Borusyak et al. (2021), this would require a firm fixed-effect control for a specification in levels.

2.4 Market Size and Innovation

We first show that our constructed export demand shock has a strong and contemporaneous impact on a firm's market size. We thus run our estimating equation (2.2) using the growth rate of sales and employment as our outcome variable $\Delta Y_{f,t}$ on the left-hand-side. We compute the average growth rate $\Delta Y_{f,t} = (Y_{f,t} - Y_{f,t-1})/[.5(Y_{f,t} + Y_{f,t-1})]$ in the same way that we constructed the export demand shock $\Delta D_{f,t}$.[17] The results for our key estimated coefficients α_τ (large darker dot) and their confidence intervals (95% as bar and 99% as dots) are represented graphically in Figure 2.3 for $\tau = -4, \ldots, 5$. The α_τ coefficients for $\tau > 0$ represent a response of the outcome variable $\Delta Y_{f,t}$ to a demand shock $\Delta_{f,t-\tau}$ τ years earlier; and conversely the coefficients for $\tau < 0$ represent a response of the outcome variable to a demand shock $-\tau$ years later.[18] It clearly shows a strong and contemporaneous response in both sales and employment to the export demand shock. As one would expect, the contemporaneous $(\tau = 0)$ employment elasticity is lower than the one for sales; but it nevertheless becomes strongly positive (and significant beyond the 1% level) in the same time interval as the demand shock. This highlights that this shock induces "real" growth for the firm (and that the increase in sales is not just associated with higher prices). As is also expected given the sluggish nature of employment adjustments, the response is longer-lasting than the one for sales and still significant one year following the demand shock. None of the pre-trend coefficients $(\tau < 0)$ are significant except for the response of sales one year prior to the demand shock. This is entirely explained by the reporting lag between the booking of an order (when it shows up in the firm's sales accounting data) and the delivery of the exported goods (when it shows up in the export customs data) – that can potentially occur in different calendar years.[19]

We now investigate how the firm's innovation responds to the same export demand shock using the same estimation strategy. We are left with a choice of functional form for a firm's patent response $\Delta Y_{f,t}$ between t and $t-1$. We do not think that the growth rate of a firm's full (over time) patent stock $P_{f,t}$ would be appropriate – because this puts too much weight on patents that may have been accumulated very far in the past and may not be relevant for more recent patents (reflecting current innovation success). Instead of dividing the change in patent stock – new patents introduced between t and $t-1$ – by the average stock in those 2 periods (the Davis-Haltiwanger growth rate), we directly control for the average number of new patent introductions $\Delta P_{f,0}$ during our pre-sample time interval

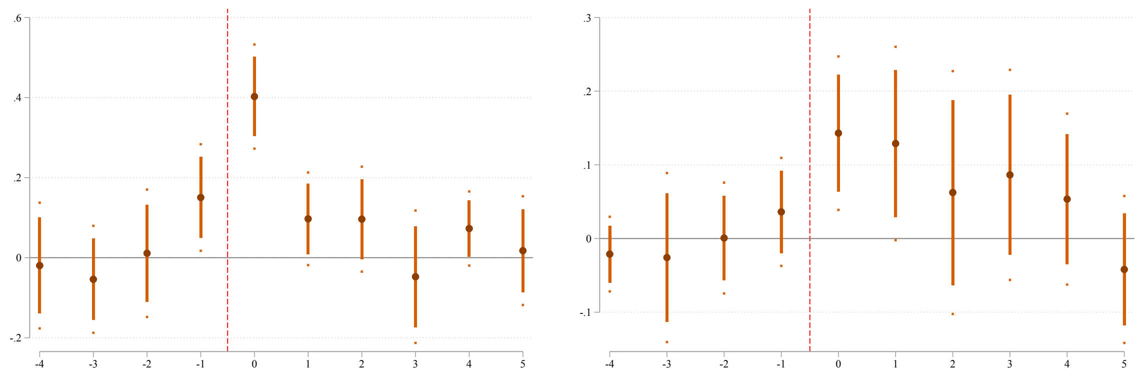
[17] Using this average growth rate computation is important for the trade shock in order to accommodate the substantial number of import flow changes to/from zero. It is inconsequential for our measurement of the growth rate of sales and employment: our results are nearly identical when we compute the growth rate using the log difference instead.

[18] From here on out, we set this timing window for the demand shock $\Delta D_{f,t}$ to 4 leads and 5 lags. We have experimented with longer and shorter windows; this does not qualitatively affect our results. See Figures 2.21 for a longer window and 2.22 for a semi-dynamic specification without pre-trends.

[19] In Appendix 2.B, we use the monthly customs export data to show that this discrepancy is explained by shipments that arrive at the beginning of a new calendar year. It also mostly affects firms with volatile sales: the significant pre-trend coefficient disappears when we exclude those firms with sales growth rates above $\pm 50\%$.

Figure 2.3: OLS: AVERAGE RESPONSE TO A DEMAND SHOCK

(a) Sales (b) Employment



Notes: Estimates of coefficients α_τ for $\tau = -4 \ldots 5$ from equation (2.2) are reported graphically with the growth rate of sales (left-hand panel) and employment (right-hand panel) as the dependent variable. The x-axis represents the value of τ, the darker dots the point estimates of α_τ, the bar the 95% confidence intervals and the smaller dots the boundaries of the 99% confidence intervals. These estimations are obtained from an OLS regression with standard errors clustered at the NACE 2-digit sector level and robust to heteroskedasticity. Number of observations: 21,421. Time period for t: 2000-2008.

from 1985-1994 (prior to t_0). Given the very large dispersion across firms in new patents $\Delta P_{f,t}$ filed during year t, including the prevalence of zeros in many years (and for many firms, most years), we use the functional form $\log(1+\Delta P_{f,t})$ with $\log(1+\Delta P_{f,0})$ in our control vector $\boldsymbol{Z}_{f,t_0}$ for our OLS specification (2.2). We also address the zeros and over-dispersion in $\Delta P_{f,t}$ using a negative binomial specification where we can then use $\Delta P_{f,t}$ directly on the left-hand-side:[20]

$$\mathbb{E}_{\boldsymbol{Z}}\left[\Delta P_{f,t}\right] = \exp\left[\boldsymbol{\alpha} \cdot \Delta_k \boldsymbol{D}_{f,t} + \boldsymbol{\gamma} \cdot \boldsymbol{Z}_{f,t_0} + \tilde{\boldsymbol{\gamma}} \cdot \left(\tilde{\boldsymbol{Z}}_{f,t_0} \times \chi_t\right)\right], \tag{2.3}$$

where the expectation $\mathbb{E}_{\boldsymbol{Z}}$ is taken conditional on $\boldsymbol{Z}_{f,t}$ and on past and future values of $\Delta D_{f,t}$. We keep the same functional form $\log(1+\Delta P_{f,0})$ in $\boldsymbol{Z}_{f,t_0}$ to control for the average rate of new patent introductions during our pre-sample years.[21] We choose a negative binomial (NB) specification as it is best suited (especially compared to Poisson) for the over-dispersion in the empirical distribution of new patents $\Delta P_{f,t}$, which standard deviation is 10.9, an order of magnitude higher than the 0.9 mean.

The graphical results for our OLS specification with the $\log(1+\Delta P_{f,t})$ functional form are presented in Figure 2.4 with the innovation response $\Delta P_{f,t}$ measured both as new priority patents as well

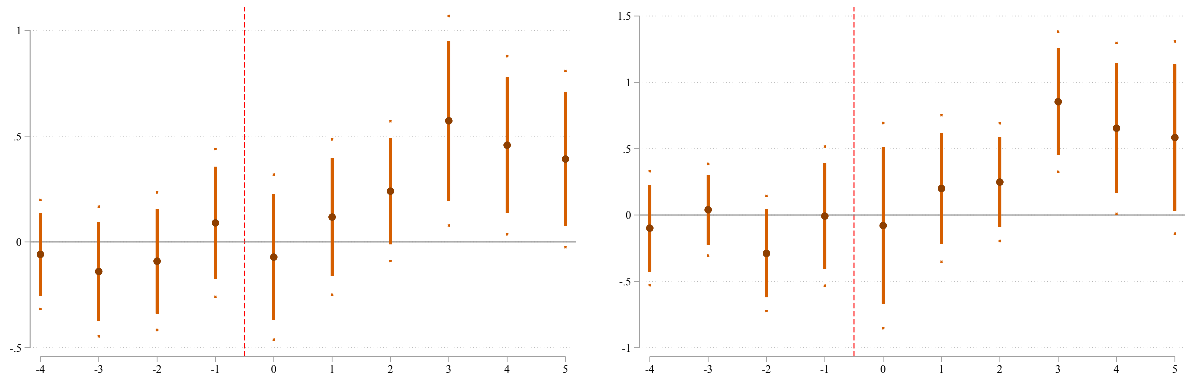
[20] In Figure 2.16 in the Appendix 2.C, we also show results using the inverse hyperbolic sine transformation of the dependent variable: $\log\left(y + \sqrt{1+y^2}\right)$.

[21] This control is then defined for firms with zero new patents during some pre-sample years. We have also experimented with using $\log \Delta P_{f,0}$ directly in $\boldsymbol{Z}_{f,t}$ – hence a control for $\Delta P_{f,0}$ outside of the exponential in (2.3) – along with an indicator variable when $P_{f,0}$ is zero. This does not qualitatively affect our results. See Blundell et al. (1999) and Aghion et al. (2016) for a use of this type of control function in a similar specification.

as our alternative measure based on citations received within five years. The graphical results for our negative binomial specification (2.3) are presented in Figure 2.5 with the same two options for the innovation response $\Delta P_{f,t}$.

Figure 2.4: OLS: AVERAGE RESPONSE TO A DEMAND SHOCK

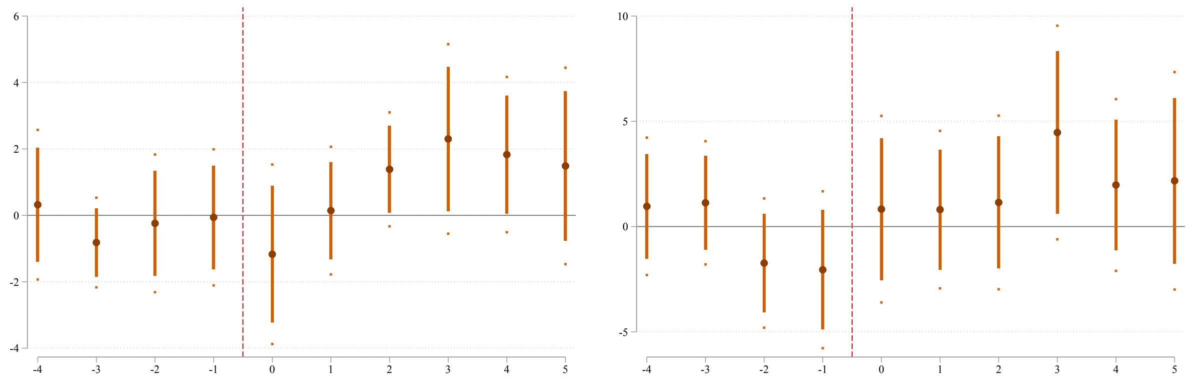
(a) Priority patents (b) Citations within 5 years



Notes: Estimates of coefficients α_τ for $\tau = -4 \ldots 5$ from equation (2.2) are reported graphically. The two panels differ in the dependent variable: the left-hand side panel considers the log of the number of new priority patents + 1 and the right-hand side panel considers the log of the number of accumulated citations received within 5 years + 1. The x-axis represents the value of τ, the darker dots the point estimates of α_τ, the bar the 95% confidence intervals and the smaller dots the boundaries of the 99% confidence intervals. These estimations are obtained from an OLS regression with standard errors clustered at the NACE 2-digit sector level and robust to heteroskedasticity. Number of observations: 22,175. Time period for t: 2000-2008.

Figure 2.5: NEGATIVE BINOMIAL: AVERAGE RESPONSE TO A DEMAND SHOCK

(a) Priority patents (b) Citations within 5 years



Notes: Estimates of coefficients α_τ for $\tau = -4 \ldots 5$ from equation (2.3) are reported graphically. The two panels differ in the dependent variable: the left-hand side panel considers the number of new priority patents and the right-hand side panel considers the number of accumulated citations received within 5 years. The x-axis represents the value of τ, the darker dots the point estimates of α_τ, the bar the 95% confidence intervals and the smaller dots the boundaries of the 99% confidence intervals. These estimations are obtained from a negative binomial regression with standard errors clustered at the NACE 2-digit sector level and robust to heteroskedasticity. Number of observations: 22,175. Time period for t: 2000-2008.

All four figures (across different functional form specifications and new patent measures) show a significant and sustained response of patenting activity starting 3 years after the export shock. The pre-trends are centered around zero and do not show any sign that the patenting activity precedes the change in export demand. We thus find a significant aggregate *market size* effect of export demand shocks on French firms' innovation. Since our specifications include sector-year fixed effects, this innovation response cannot be explained by any sector-wide innovation changes. Rather, it must be driven by the firm-level innovation responses to demand shocks.

Table 2.6 summarizes our results from Figures 2.3-2.5 for the response of both market size (scale) and innovation to the export demand shock. The dynamic leads and lags are cumulated (the coefficients are summed) into a pre-period (1 to 4 years prior to the shock), a current period (concurrent and 1 year after the shock), and a future period (2 to 5 years after the shock). Even when cumulated, there is no evidence of pre-trends for either scale (sales and employment) or innovation. Table 2.6 also highlights how the response of scale occurs concurrently with the shock while the response of innovation is delayed to the future period. This cumulative response is significant beyond the 1% level in our OLS specifications, and significant around the 5% level (a bit stronger for the patents; and weaker when measured as citations) in our negative binomial specification.[22]

The economic magnitude of those cumulated innovation responses are substantial.[23] On average, there are 3,321 firms (the innovators in our sample) operating in the future period 2-5 years following a demand shock in 1999, 2000, ..., 2003. Those firms introduced 8,176 priority patents (on average, in that same future period), which generated 27,982 citations. The future period coefficients for innovation in Table 2.6 imply that a 1 point export demand shock would induce 52 (OLS) - 166 (NB) new priority patent associated with 82 (OLS) - 1,344 (NB) citations during that same future period (again, on average for demand shocks in 1999, 2000, ... , 2003.) This represents an aggregate (macro) elasticity of .64-2.0 for patents to an aggregate export demand shock; and an elasticity of .29-4.8 in terms of citations. The economic magnitude of that innovation response to demand shocks in export markets is therefore substantial.[24]

[22] As the discussion of the economic magnitudes below makes clear, this is due to very large but imprecisely estimated coefficients in the negative binomial specification. The results from Tables 2.6 and 2.7, which use 2-digit NACE controls, are robust to using 5-digit NACE controls.

[23] The demand shock has a mean of 0.011 with a standard deviation of 0.043. The 25^{th} percentile is -0.002 and the 75^{th} is 0.025. Its yearly mean changes in line with international trade. A one point change in the export demand shock corresponds to a fourth of a standard deviation and is a small move in the distribution (moving from the bottom 10% to the top 90% corresponds to an increase of 9pp).

[24] We have chosen throughout to report the magnitudes of the innovation responses in terms of the export demand shock. We could alternatively consider an instrumental variable specification in order to report those innovation magnitudes in terms of a shock to scale (market size or employment), using our scale regression as a first stage. Our innovation regressions can be viewed as the reduced form for that instrumental variable specification. Since the magnitude of those reduced form coefficients have a natural and direct interpretation, we stick to this specification.

Table 2.6: CUMULATIVE RESPONSE TO DEMAND SHOCK

	Scale		Innovation			
	Sales	Employment	Priority Patents		Citations	
	OLS	OLS	OLS	NB	OLS	NB
Pre-Trend	0.088 (0.134)	-0.010 (0.062)	-0.199 (0.328)	-0.799 (1.588)	-0.357 (0.448)	-1.696 (2.755)
Current	0.500*** (0.073)	0.272*** (0.065)	0.046 (0.257)	-1.029 (1.580)	0.121 (0.474)	1.632 (2.874)
Future	0.139 (0.094)	0.160 (0.113)	1.662*** (0.534)	7.002** (3.112)	2.341*** (0.718)	9.765* (5.262)

Notes: This table reports point estimate and standard errors (under parentheses) for different linear combinations of coefficients from various estimations of equations (2.2) and (2.3). Pre-Trend corresponds to the estimate of $\alpha_{-4}+\alpha_{-3}+\alpha_{-2}+\alpha_{-1}$, current to $\alpha_0+\alpha_1$ and Future to $\alpha_2+\alpha_3+\alpha_4+\alpha_5$. Column 1 corresponds to the results displayed in Figure 2.3a, column 2 to Figure 2.3b, column 3 to Figure 2.4a, column 4 to Figure 2.5a, column 5 to Figure 2.4b and column 6 to Figure 2.5b. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

2.5 Heterogeneous Impact: Distance to Frontier

We now investigate whether this innovation response varies across firms based on their distance to their sector's frontier. We use labor productivity (value-added per worker) as our metric for this distance. Just as we did with the firm-level export shares, we use the initial year t_0 to generate a distance measure that does not subsequently vary over time $t > t_0$. We partition firms into those with productivity above their 2-digit sector median (in year t_0), $a_{f,t_0} \geq \overline{a}_{t_0}$ (represented by indicator dummy $\mathbf{1}_a^+$), and those with productivity below the sector median, $a_{f,t_0} < \overline{a}_{t_0}$ (represented by indicator dummy $\mathbf{1}_a^-$). More specifically, we consider the following regression equation:

$$\Delta Y_{f,t} = \boldsymbol{\alpha}_H \cdot \left(\Delta_k \boldsymbol{D}_{f,t} \times \mathbf{1}_a^+\right) + \boldsymbol{\alpha}_L \cdot \left(\Delta_k \boldsymbol{D}_{f,t} \times \mathbf{1}_a^-\right) + \boldsymbol{\gamma} \cdot \boldsymbol{Z}_{f,t_0} + \tilde{\boldsymbol{\gamma}} \cdot \left(\tilde{\boldsymbol{Z}}_{f,t_0} \times \chi_t\right) + \varepsilon_{f,t}. \quad (2.4)$$

Since the firm's initial productivity level a_{f,t_0} is now used to construct our two different trade shocks on the right-hand-side, we add that variable to the control vectors $\boldsymbol{Z}_{f,t_0}$ and $\tilde{\boldsymbol{Z}}_{f,t_0}$. We use the same functional form $\Delta Y_{f,t} = \log(1 + \Delta P_{f,t})$ for our OLS specification (adding $\log(1 + \Delta P_{f,0})$ to our control vector $\boldsymbol{Z}_{f,t_0}$). And we also estimate a negative binomial specification with the 'untransformed' new patent measure $\Delta P_{f,t}$ on the left-hand side, along with a control for $\Delta P_{f,0}$ in $\boldsymbol{Z}_{f,t_0}$:

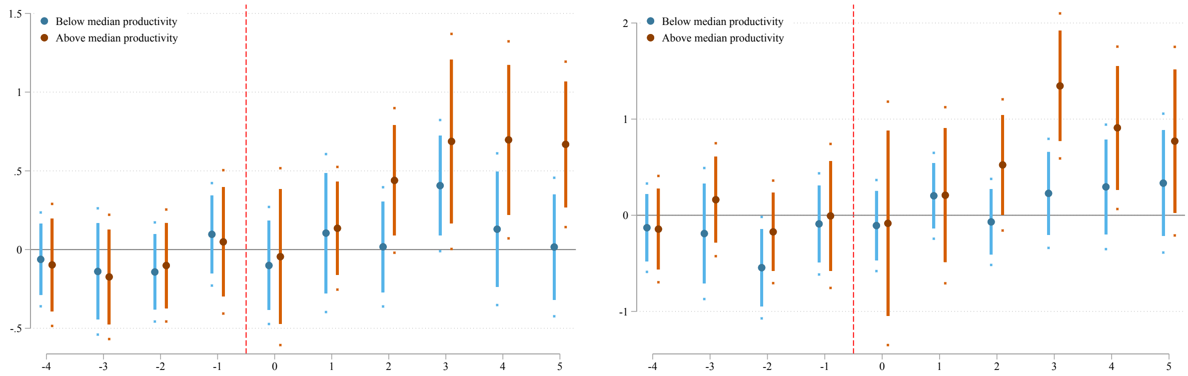
$$\mathbb{E}_{\boldsymbol{Z}}\left[\Delta P_{f,t}\right] = \exp\left[\boldsymbol{\alpha}_H \cdot (\Delta_k \boldsymbol{D}_{f,t} \times \mathbf{1}_a^+) + \boldsymbol{\alpha}_L \cdot (\Delta_k \boldsymbol{D}_{f,t} \times \mathbf{1}_a^-) + \boldsymbol{\gamma} \cdot \boldsymbol{Z}_{f,t_0} + \tilde{\boldsymbol{\gamma}} \cdot \left(\tilde{\boldsymbol{Z}}_{f,t_0} \times \chi_t\right)\right], \quad (2.5)$$

where the expectation $\mathbb{E}_{\boldsymbol{Z}}$ is again taken conditional on $\boldsymbol{Z}_{f,t}$ and on past and future values of $\Delta D_{f,t}$.

The graphical results for both our OLS and negative binomial specifications are presented in Figures 2.6 and 2.7, once again using both priority patents and the accumulated citations as our measure of new patent activity $\Delta P_{f,t}$. All four figures show a significant and sustained response of patenting activity starting 3 years after the export shock – but *only* for firms that are initially closer to their sector's frontier (with labor productivity above the median level, in orange). In Appendix 2.C, we return to the full battery of robustness checks that we previously described for the analysis of the non-heterogeneous responses. The main messages from Figures 2.6 and 2.7 remain unchanged (See Figures 2.15-2.22).

Figure 2.6: OLS: HETEROGENEOUS RESPONSE TO A DEMAND SHOCK

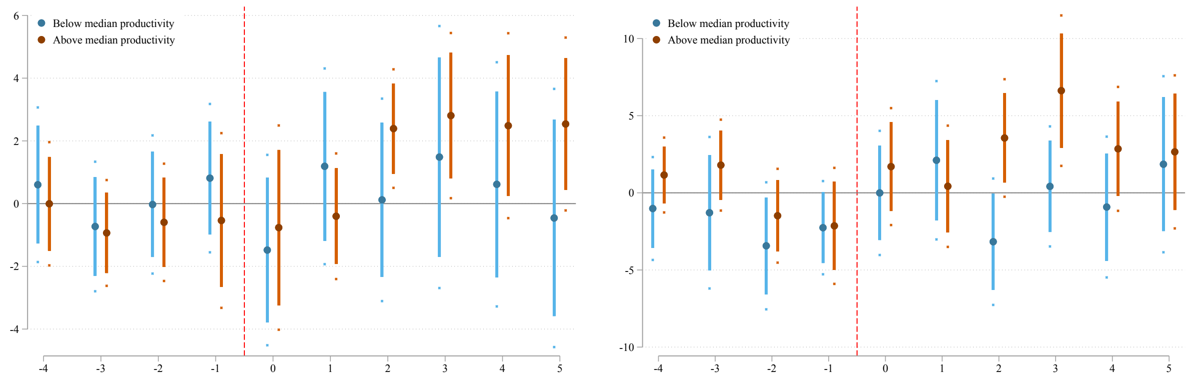
(a) Priority Patents (b) Citations within 5 years



Notes: Estimates of coefficients $\alpha_{H,\tau}$ and $\alpha_{L,\tau}$ for $\tau = -4 \ldots 5$ from equation (2.4) are presented graphically, respectively in orange and blue. The two panels differ in the dependent variable: the left-hand side panel considers the log of the number of new priority patents + 1 and the right-hand side panel considers the log of the number of accumulated citations received within 5 years + 1. The x-axis represents the value of τ, the darker dots the point estimates of α_τ, the bar the 95% confidence intervals and the smaller dots the boundaries of the 99% confidence intervals. These estimations are obtained from an OLS regression with standard errors clustered at the NACE 2-digit sector level and robust to heteroskedasticity. Number of observations: 22,175. Time period for t: 2000-2008.

Figure 2.7: NEGATIVE BINOMIAL: HETEROGENEOUS RESPONSE TO A DEMAND SHOCK

(a) Priority Patents (b) Citations within 5 years

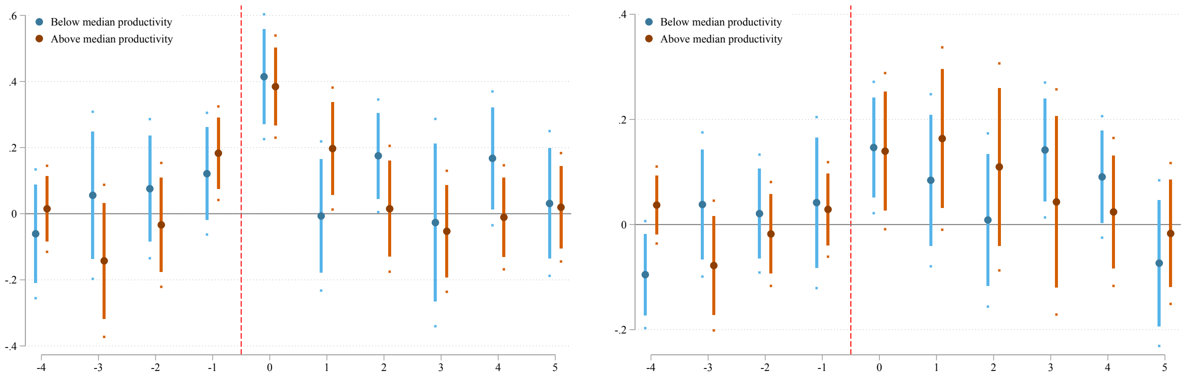


Notes: Estimates of coefficients $\alpha_{H,\tau}$ and $\alpha_{L,\tau}$ for $\tau = -4 \ldots 5$ from equation (2.5) are presented graphically, respectively in orange and blue. The two panels differ in the dependent variable: the left-hand side panel considers the number of new priority patents and the right-hand side panel considers the number of accumulated citations received within 5 years. The x-axis represents the value of τ, the darker dots the point estimates of α_τ, the bar the 95% confidence intervals and the smaller dots the boundaries of the 99% confidence intervals. These estimations are obtained from a negative binomial regression with standard errors clustered at the NACE 2-digit sector times productivity group level and robust to heteroskedasticity. Number of observations: 22,175. Time period for t: 2000-2008.

Could this heterogeneous response simply reflect the fact that the demand shock only affects the most productive firms? To check that this is not the case, we replicate the results shown in Figures 2.3a and 2.3b: that is, we allow for a different impact of the export demand shocks on sales or employment depending upon initial productivity levels. Looking at Figures 2.8a and 2.8b, we see that in contrast to what we observe for innovation, there is no heterogeneous response of sales or employment to a demand shock for low versus high productivity firms. The responses for both sets of firms match the magnitudes of the average response that we previously documented.[25]

Figure 2.8: HETEROGENEOUS RESPONSE TO A DEMAND SHOCK - SALES AND EMPLOYMENT

(a) Sales (b) Employment



Notes: This Figure replicates Figure 2.3 but allowing for heterogeneity based on the initial productivity level as described in equation (2.4). Number of observations: 21,421. Time period for t: 2000-2008.

We summarize once again our dynamic results in Figures 2.6-2.8 by cumulating the coefficients into pre-trend, current, and future periods just as we previously reported in Table 2.6 for the case without the heterogeneous impact by productivity: pre-trend for 1-4 years prior to the export demand shock; current for 0-1 year following the shock; and future for 2-5 years following the shock. Those coefficients are reported in Table 2.7 for the above and below median firm productivity groups. In addition, we now report a significance test for their difference across those two groups.

As can also be seen in Figures 2.6b and 2.7b, the response of innovation in terms of citations is negative in the pre-trend period for below-median productivity firms. There is no evidence of pre-trends in any of the other dependent variables for either scale of innovation among either subset of more or less productive firms. Table 2.7 also highlights the strong contemporaneous response of the scale variables to the demand shock that we previously emphasized. Importantly, there is no evidence of any significant differences in the responses of those scale variables across the two productivity groups. On the other hand, Table 2.7 makes clear that the strong future response of innovation for the above-median firms is statistically distinguishable from the response for the below-median firms. Not only are the future response coefficients for those below-median firms insignificant, the coefficient difference in favor of the relatively more productive firms is statistically significant beyond the 1%

[25] As can be seen in Figure 2.8a, the growth rate of the sales response for the below median firms fluctuates up and down following the trade shock. This effect is driven by firms with volatile sales: it disappears when we exclude those firms with sales growth rates above $\pm 50\%$.

Table 2.7: CUMULATIVE HETEROGENEOUS RESPONSE TO DEMAND SHOCK

	Scale		Innovation			
	Sales	Employment	Priority Patents		Citations	
	OLS	OLS	OLS	NB	OLS	NB
Pre-Trend						
Below Median	0.192 (0.152)	0.006 (0.080)	-0.247 (0.328)	0.658 (2.016)	-0.954* (0.484)	-7.999** (3.662)
Above Median	0.022 (0.175)	-0.030 (0.087)	-0.323 (0.381)	-2.074 (1.619)	-0.162 (0.665)	-0.675 (2.741)
Difference	-0.170 (0.200)	-0.036 (0.087)	-0.077 (0.433)	-2.732 (2.537)	0.792 (0.694)	7.325* (4.025)
Current						
Below Median	0.408*** (0.108)	0.231*** (0.076)	0.004 (0.251)	-0.292 (1.401)	0.096 (0.299)	2.105 (2.753)
Above Median	0.582*** (0.081)	0.303*** (0.105)	0.090 (0.300)	-1.167 (1.824)	0.124 (0.783)	2.126 (2.701)
Difference	0.174 (0.117)	0.072 (0.119)	0.087 (0.272)	-0.875 (1.523)	0.028 (0.701)	0.020 (2.549)
Future						
Below Median	0.347** (0.149)	0.168* (0.097)	0.569 (0.353)	1.761 (3.046)	0.788 (0.681)	-1.820 (3.788)
Above Median	-0.030 (0.156)	0.160 (0.146)	2.491*** (0.713)	10.223*** (3.004)	3.550*** (0.881)	15.685*** (4.880)
Difference	-0.377* (0.205)	-0.008 (0.157)	1.922*** (0.589)	8.461*** (2.725)	2.762*** (0.954)	17.505*** (3.375)

Notes: This table reports point estimate and standard errors (under parentheses) for different linear combinations of coefficients from various estimations of equations (2.4) and (2.5). Pre-Trend corresponds to the estimate of $\alpha_{X,-4}+\alpha_{X,-3}+\alpha_{X,-2}+\alpha_{X,-1}$, current to $\alpha_{X,0}+\alpha_{X,1}$ and Future to $\alpha_{X,2}+\alpha_{X,3}+\alpha_{X,4}+\alpha_{X,5}$ where $X=H$ for lines “Above Median” and $X=L$ for lines "Below Median". The lines “Difference” corresponds to the difference between the corresponding above and below median linear combinations. Column 1 corresponds to the results displayed in Figure 2.8a, column 2 to Figure 2.8b, column 3 to Figure 2.6a, column 4 to Figure 2.7a, column 5 to Figure 2.6b and column 6 to Figure 2.7b. ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively.

level in all of our specifications. This strongly supports our main finding that the innovation response is entirely concentrated within the subset of relatively more productive firms.

The economic magnitude of the innovation response for those above-median firms corresponds roughly to a similar aggregate response as the one we reported for the case without firm heterogeneity – except that this response is now concentrated more intensely and exclusively within the top-half of relatively more productive firms. The OLS coefficients in Table 2.7 imply that an aggregate 1 point increase in the export demand shock would induce 44 patents associated with 82 citations amongst the above-median firms. Those numbers are slightly lower than the 52 patents and 88 citations we previously recorded for the aggregate response without firm heterogeneity. The NB coefficient for the patent response implies a slightly higher number of patents from the 1 point increase in the export demand shock: 197 patents relative to 166 for the case without firm heterogeneity. The NB coefficient for the citation response implies a substantially larger response relative to the case without heterogeneity (almost double). But both NB coefficients for citations (and especially the one for the above-median firms) have a large standard error, so there is still a wide overlap between our predictions for the aggregate response in terms of citations with and without firm heterogeneity.

Once again, we find that the economic magnitude of the innovation response – concentrated within the subset of relatively more productive firms – is substantial.

2.6 Falsification Tests

In order to reinforce our finding of a causal impact for market size (via our demand shocks) on innovation by above-median productivity firms, we develop a falsification test that highlights that those innovation responses cannot be explained by a firm-level trend: that is, that those firms observed to increase innovation would have done so anyway absent an increase in export demand.[26] This test also provides a further check on our control that the innovation response is not explained by a firm's prior exposure to export markets (since we use prior export intensity to construct our trade shocks).[27] Our test involves the construction of placebo demand shocks for each firm and then showing that firm innovation does not respond to this placebo shock.

In our first placebo construction, we allocate products to firms randomly (based on their empirical distribution across firms) and compute the demand shocks that each firm would have experienced had it actually exported those products at t_0. In our second placebo construction, we instead allocate the export destinations randomly across firms (again, based on the empirical distribution of destinations across firms).[28]

We construct 2000 different placebo demand shocks using both methods, and then estimate our baseline OLS specification (2.4) each time with the response of priority patents on the left-hand side. Figure 2.9 shows the cumulative distribution for the coefficient $\alpha_{H,4}$ and its t-statistic for the response by firms with above-median productivity 4 years after the shock. Against those distributions, we show (red vertical line) the coefficient value and t-statistic for $\alpha_{H,4}$ that we reported in Figure 2.6a using the 'true' demand shocks. We immediately see that the value and significance of the demand shock coefficient we previously obtained are clear outliers in those distributions (well beyond the 100th percentile for the coefficient values; and at the 98.5 and 96 percentiles for the associated t-statistics). We can thus easily reject the hypothesis that a similar innovation response by the above-median productivity firms would have been observed absent the impact of the "true" demand shock. We have repeated this falsification test summing the coefficients representing 2 to 5 years after the shock (instead of just year 4), along with its associated t-statistic. In all those cases, our reported coefficients (and their t-statistics) are again clear outliers in the simulated cdf: above the 95th percentile of the distribution in all cases (and above the 100th percentile in a few). We have also repeated this exercise, with similar results, with citations as the dependent variable.

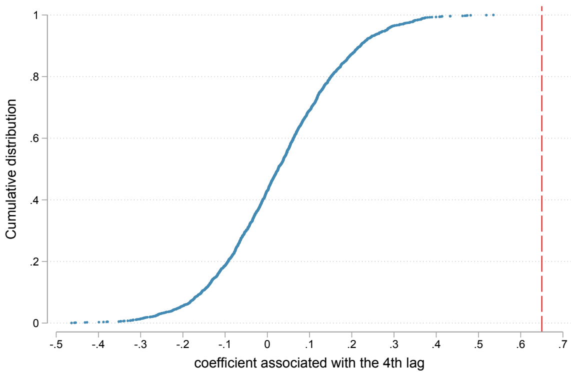
[26]A similar approach has been implemented by Chetty et al. (2009) and Malgouyres et al. (2019).

[27]Our main check is to add export intensity interacted with the year fixed effects as controls.

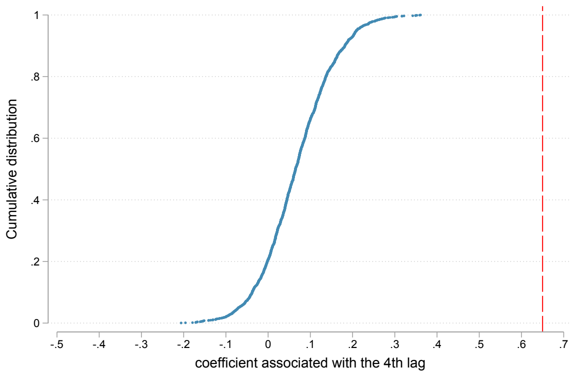
[28]To be more precise, each placebo demand shock is the outcome of a random permutation across firms from either the empirical distribution of products, or the empirical distribution of destinations.

Figure 2.9: OLS: FALSIFICATION TESTS

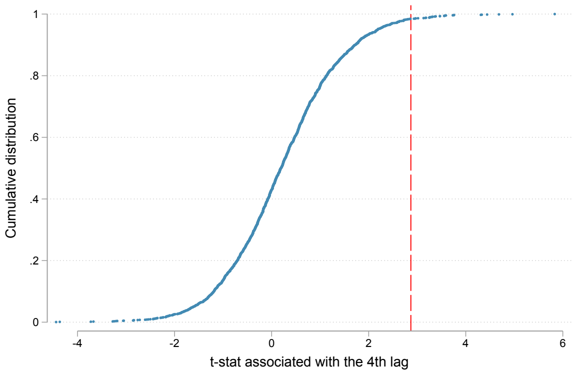
(a) Randomly Switch Products



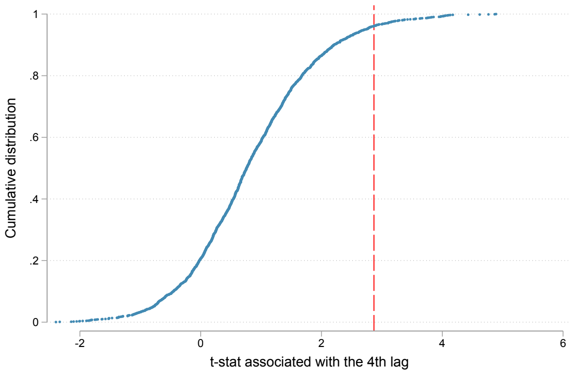
(b) Randomly Switch Destinations



(c) Randomly Switch Products



(d) Randomly Switch Destinations



Notes: This figure plots the cumulative distribution of the point estimates (top panels) and the associated t-stat (bottom panels) and for the $\alpha_{H,4}$ coefficient when equation (2.4) is estimated 2000 times with a placebo shock, randomly switching the products exported at t_0 (left panel) or randomly switching the export countries at t_0 (right panel). $\alpha_{H,4}$ coefficient and t-stat from Figure 2.6a in red line.

2.7 A model

2.7.1 Presentation

In this section, we show that our main finding of a skewed innovation response to common demand shocks arises naturally from a model of endogenous innovation and competition with firm heterogeneity. Our model features a "standard" market size effect that increases innovation for all firms. But it also embodies an endogenous competition effect that discourages innovation by low productivity firms. This skewed induced competition effect captures the idea that the expanded market for exports will attract new firms into the export market as more firms find it profitable to sell their products there; this in turn will raise competition for exporters into that market. Due to the nature of competition between firms – featuring endogenous markups – this effect gradually dissipates as productivity (and resulting market share) increases. This competition effect is thus more salient for smaller French firms with initially lower productivity, as they lose market share to larger more productive firms.

The model we present is highly parametrized. However, we show in a companion paper (Aghion et al., 2018) that an increase in market size triggers a skewed competition effect under more general cost (including the return to innovation) and demand conditions. In particular, we show that the main skewness result holds for a broad class of preferences under monopolistic competition that satisfy Marshall's Second Law of Demand (MSLD), i.e. lead to residual (firm-level) inverse demands that become more inelastic as consumption increases. Instead, a model with monopolistic competition and CES preferences (and hence exogenous markups) would not generate a skewed induced competition effect of increased market size. The recent empirical trade literature provides mounting evidence for the relevance of endogenous markups associated with MSLD demand.[29]

Finally, we stress that our empirical work and results in the previous sections are not meant to specifically test whether the heterogeneous impact of increased market size on innovation is due to the skewed competition effect with endogenous markups that we model in this section. We are just showing that this evidence is consistent with – and easily explained by – a competition channel highlighted by our model. Our model also illustrates the fact that very few assumptions are needed beyond MSLD demand to generate a skewed innovation response to increased market size.

[29] See Melitz (2018) for a summary of this evidence and how it is connected to endogenous markups and MSLD demand. This evidence for endogenous markups adjustments would also be consistent with oligopoly models where the elasticity of substitution between products remain constant. Such a model would nevertheless feature endogenous price elasticities that respond in a very similar way to those in a model of monopolistic competition with MSLD demand.

2.7.2 Basic setup

French firms exporting to some export market destination D are competing with local firms producing in D. We let L denote the number of consumers in that destination. This indexes market size. These consumers have preferences over all varieties available in D. There is a continuum of differentiated varieties indexed by $i \in [0, M]$, where M is the measure of available products. Suppose that the demand for variety q_i is generated by a representative consumer in country D with additively separable preferences with sub-utility:[30]

$$u(q_i) = \alpha q_i - \frac{\beta q_i^2}{2},$$

where $\alpha > 0$ and $\beta > 0$.

Those preferences do not differentiate between French or locally produced varieties. Thus, the output, profit and revenues for the French exporters and local producers have the same expression. For simplicity, we assume that both types of firms have access to the same innovation technology, which leads to similar innovation decisions.

Consumer optimization

This representative consumer facing prices p_i solves:

$$\max_{q_i \geq 0} \int_0^M u(q_i)di \text{ s.t. } \int_0^M p_i q_i di = 1.$$

This yields the inverse residual demand function (per consumer):

$$p(q_i) = \frac{u'(q_i)}{\lambda} = \frac{\alpha - \beta q_i}{\lambda}, \tag{2.6}$$

where $\lambda = \displaystyle\int_0^M u'(q_i) q_i di > 0$ is the corresponding Lagrange multiplier, also equal to the marginal utility of income. Given the assumption of separable preferences, this marginal utility of income λ is the unique endogenous aggregate demand shifter. Higher λ shifts all residual demand curves downwards; we thus interpret this as an increase in competition for a given exogenous level of market size L.

[30] As we previously discussed, our analysis can be extended to a broader class of preferences that satisfy Marshall's Second Law of Demand (such that residual demand becomes more inelastic as consumption increases).

Firm optimization

Consider a (French or domestic) firm with marginal cost c facing competition λ. This firm chooses the output per consumer $q(c;\lambda)$ to maximize operating profits $L\left[p(q)q - cq\right]$. The corresponding first order condition yields

$$q(c;\lambda) = \frac{\alpha - c\lambda}{2\beta}, \tag{2.7}$$

so long as the firm's cost is below α/λ; the remaining firms with higher cost do not produce. This output choice in turn leads to the maximized profit per consumer

$$\pi(c;\lambda) = \frac{(\alpha - c\lambda)^2}{4\beta\lambda}.$$

In particular, we see that both output and profit are decreasing in both firm level cost c and the endogenous competition measure λ. More productive firms (with lower cost c) are larger and earn higher profits than their less productive counterparts; and an increase in competition λ lowers production levels and profits for all firms.

Innovation choice

A firm is characterized by its baseline cost $\tilde{c}$. It can reduce its marginal cost of production c below its baseline cost by investing in innovation. More formally, we assume that

$$c = \tilde{c} - \varepsilon k,$$

where k is the firm's investment in innovation and $\varepsilon > 0$; and we assume that the cost of innovation is quadratic in k, equal to $c_I k + \frac{1}{2} c_{I2} k^2$.[31]

Thus a firm with baseline cost $\tilde{c}$ will choose its optimal R&D investment $k(\tilde{c};\lambda)$ so as to maximize total profit:

$$\Pi(\tilde{c}, k;\lambda) = L\pi(\tilde{c} - \varepsilon k;\lambda) - c_I k - \frac{1}{2} c_{I2} k^2.$$

The optimal R&D investment $k(\tilde{c};\lambda)$, if positive, satisfies the first order condition:

$$\varepsilon Q(\tilde{c}, k;\lambda) = c_{I2} k + c_I, \tag{FOC}$$

[31] Since we only consider a single sale destination D for our firms, we are implicitly assuming that the innovation is directed at the delivered cost to consumers in D. We should thus think of innovation as specific to the appeal/cost trade-off to consumers in D. Our companion paper describes how our main skewness result holds for more general functional forms for the cost and return to innovation.

where

$$Q(\tilde{c}, k; \lambda) \equiv Lq(\tilde{c} - \varepsilon k; \lambda) = L[\alpha - (\tilde{c} - \varepsilon k)\lambda]/2\beta$$

is the total firm output (across consumers) produced by a firm with baseline cost $\tilde{c}$ and innovation k. We assume that the baseline cost $\tilde{c}$ is bounded below by $\tilde{c}_{\min}$ such that $\tilde{c}_{\min} - \varepsilon k(\tilde{c}_{\min}; \lambda) = 0$, or equivalently

$$\tilde{c}_{\min} = \frac{\varepsilon}{c_{I2}} \left(\frac{\varepsilon L \alpha}{2\beta} - c_I \right).$$

This in turn ensures that the post-innovation marginal cost is bounded away from zero, even for the most productive firms.

Figure 2.10 depicts the optimal innovation choice at the intersection between the marginal cost (*MC*, right-hand side of FOC) and the marginal benefit of innovation (*MB*, left-hand side of FOC). As long as the marginal benefit is above the marginal cost of investing in R&D, the firm wants to increase innovation, because the marginal profit made by investing one more unit of R&D exceeds its marginal cost. We assume that the second order condition holds so that the slope of the marginal cost is strictly larger than the slope of the marginal gain:

$$c_{I2} > \varepsilon \frac{\partial Q}{\partial k} = \frac{\varepsilon^2 \lambda L}{2\beta}. \tag{SOC}$$

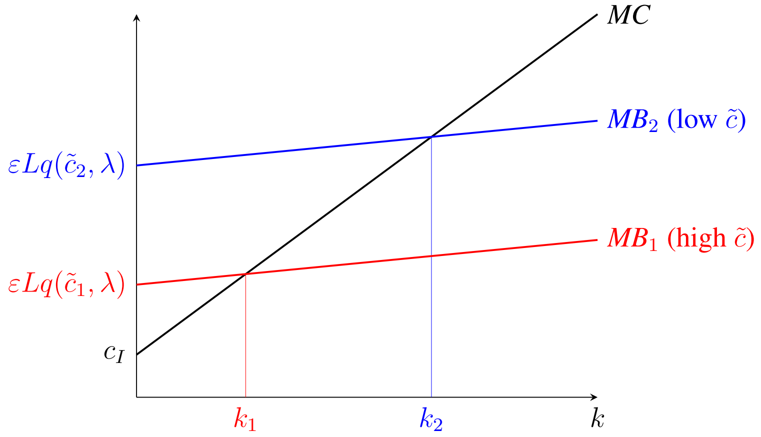
This ensures a smooth innovation response to productivity differences.

When comparing a more productive firm (with lower baseline cost, depicted by the blue curve) and a less productive firm (with higher baseline cost, depicted by the red curve), we see that both firms face the same marginal cost curve and their marginal gain curves have the same slope. Only the zero intercepts of the two marginal gain curves are different: the lower $\tilde{c}$ firms have a higher intercept, thus a higher marginal gain, and therefore invest more in R&D. Firms with sufficiently high baseline costs do not innovate, as the zero intercept of their marginal gain curves falls below c_I, so that even their first innovation unit would not be worth its cost. These are firms with baseline costs above the baseline cost of the marginal innovator, which is equal to:

$$\widehat{C}_I = \frac{1}{\lambda} \left(\alpha - \frac{2\beta c_I}{\varepsilon L} \right). \tag{2.8}$$

In the next subsection we analyze how the optimal innovation choice $k(\tilde{c}; \lambda)$ responds to a positive demand shock, i.e. to an increase in market size L.

Figure 2.10: Optimal innovation is higher for more efficient firms



2.7.3 The market size and competition effects

We first analyze the direct effect of an increase in L, holding the competition level λ constant. At each firm's current innovation choice $k(\tilde{c}; \lambda)$, this triggers a proportional increase in firm output, and an upward shift in the marginal benefit of innovation, inducing all firms to increase innovation.

Figure 2.11 shows this innovation response for firms with different baseline costs. Both the intercept and the slope of the marginal gain curve increase. We see how this leads to higher innovation for all firms. Given our assumptions on the benefits and costs of innovations, this leads to higher innovation responses for more productive firms:

$$\frac{\partial^2 k}{\partial L \partial \tilde{c}} < 0.$$

This increase in market size also induces some firms to begin R&D (higher $\widehat{C}_I$, see 2.8).

We now consider the effect of an increase in competition λ, holding market size L constant. At each firm's current innovation choice $k(\tilde{c}; \lambda)$, this triggers a decrease in firm output (see equation (2.7)). However, unlike the case of a change in market size L, this output response is no longer proportional across firms: high cost firms bear the brunt of the competition increase and disproportionately lose market share. Even though all firms respond by reducing innovation, this reduction in innovation is most pronounced (larger) for those high cost firms:

$$\frac{\partial^2 k}{\partial \lambda \partial \tilde{c}} < 0.$$

This contrasts with the case of a market size decrease (leading to proportional output decreases), which would lead to bigger innovation reductions for low cost firms instead. In the limit for the most

Figure 2.11: Direct market size effect (increase in L)

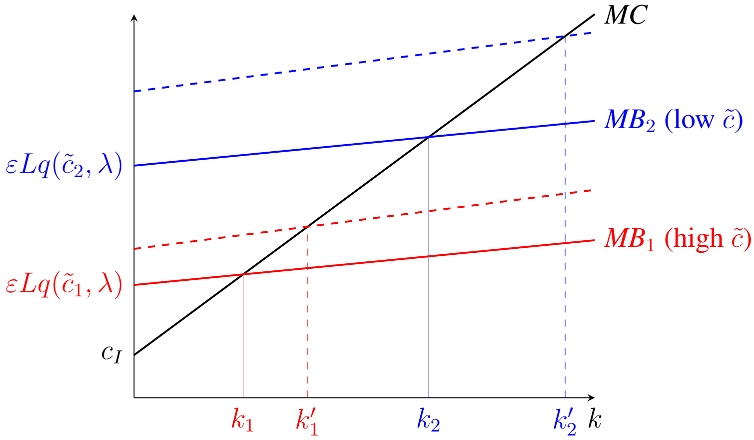
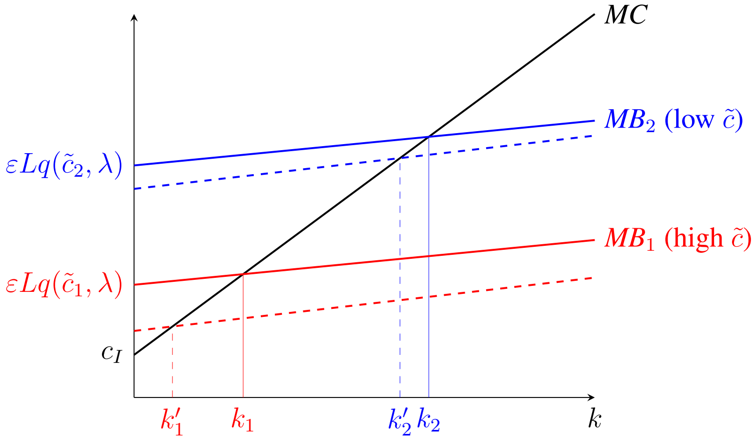


Figure 2.12: Competition effect (increase in λ)



efficient firms (with baseline cost approaching $\tilde{c}_{\min}$), the negative impact of increased competition on innovation dissipates completely (see FOC).

Figure 2.12 shows this innovation response for firms with different baseline costs. The increase in competition decreases the marginal benefit of innovation, but substantially more for the high cost firm – because the intercept decrease is larger (recall that the slope of the marginal benefit curve does not change with the firm's baseline cost).[32] Thus, the high cost firm's reduction in innovation is most pronounced. The competition increase also induces some firms to stop R&D (lower $\widehat{C}_I$, see 2.8).

[32]The new dotted marginal benefit curve remains below the old one at least until it meets the marginal cost curve, even though an increase in competition increases the slope of the marginal benefit curve.

2.7.4 The heterogeneous innovation response to an export shock

How can our model generate the skewness we observed in firms' innovation response to a positive export demand shock? In the Appendix we endogenize the equilibrium competition level λ in country D and we show that it increases with L. The intuition is that an increase in market size L induces entry on the export market D by new firms; this in turn increases the elasticity of the inverse demand curve faced by each French exporter to D and an increase in λ. It then follows that an increase in market size L will have two effects on firms' innovation incentives: (a) a *direct* - positive - *market size effect*, whereby the increase in L induces all firms to increase innovation; this effect was shown above to be more positive for more frontier firms (i.e. for firms with lower initial production cost $\tilde{c}$); (b) an *induced* - negative - *competition effect* whereby the increase in L increases competition λ which in turns reduces firms' innovation incentives; as we saw above, the effect of an increase in λ on firms' innovation is more negative for less productive firms (i.e. for firms with higher initial production cost $\tilde{c}$). The overall effect of an increase in market size L on innovation – which combines the direct market size effect and the induced competition effect – will be unambiguously more positive for more frontier firms; moreover, this overall effect can turn out to be negative for the least productive firms - depending on the relative magnitude of the direct and indirect impacts. This heterogeneous response is fully consistent with our empirical analysis: we showed that the most productive half of the firms increase their innovation when their market size expands, while the response for the least productive half of the firms is essentially muted.

2.8 Conclusion

In this paper we used exhaustive data covering the French manufacturing sector to analyze the impact of export demand shocks on patenting by French exporting firms. To disentangle the direction of causality between export demand and innovation, we constructed a firm-level export demand shock which responds to aggregate conditions in a firm's export destinations but is exogenous to firm-level decisions.

We first showed that French firms respond to exogenous growth shocks in their export destinations by patenting more. Second, we showed that this positive impact of market size on innovation is skewed and entirely driven by French firms with above-median initial labor productivity within their sector. Third, we showed that the innovation response arises 2 to 5 years after a demand shock, whereas the same demand shock raises contemporaneous sales and employment for all firms. And lastly, we developed a simple theoretical model with endogenous innovation and endogenous markups which rationalizes the skewed innovation response to increases in export demand.

Our paper contributes to the existing literature on innovation and market size in several respects: To our knowledge, we are the first to identify a causal impact of firm-level market size on innovation that is independent of any sector-level dynamics (controlling for arbitrary sector level year-on-year changes) and widespread across the entire manufacturing sector. Given the detailed timing of the changes in demand, we are also able to precisely measure the time-lag required before the ensuing patenting activity is recorded. And lastly, we have showed that this innovation response is highly skewed and dominated by relatively more productive firms within each sector.

Bibliography

Acemoglu, Daron and Joshua Linn, "Market size in innovation: theory and evidence from the pharmaceutical industry," *The Quarterly journal of economics*, 2004, *119* (3), 1049–1090.

Aghion, Philippe and Peter Howitt, *The economics of growth*, Cambridge, Massachusetts. London, England: MIT Press, 2009. OCLC: 610950320.

_ , Antoine Dechezleprêtre, David Hemous, Ralf Martin, and John Van Reenen, "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry," *Journal of Political Economy*, 2016, *124* (1), 1–51.

_ , Antonin Bergeaud, Matthieu Lequien, and Marc J Melitz, "The Impact of Exports on Innovation: Theory and Evidence," Working Paper 24600, National Bureau of Economic Research May 2018.

_ , Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: an Inverted-U Relationship," *The Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Akcigit, Ufuk, Sina T. Ates, and Giammario Impullitti, "Innovation and Trade Policy in a Globalized World," NBER Working Paper 24543, National Bureau of Economic Research 2018.

Amiti, Mary, Stephen J Redding, and David E Weinstein, "The impact of the 2018 tariffs on prices and welfare," *Journal of Economic Perspectives*, 2019, *33* (4), 187–210.

Autor, David, David Dorn, Gordon Hanson, Gary Pisano, and Pian Shu, "Foreign Competition and Domestic Innovation: Evidence from U.S. Patents," Technical Report 22879, National Bureau of Economic Research 2016. DOI: 10.3386/w22879.

_ , _ , Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," Working Paper 23396, National Bureau of Economic Research May 2017.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?* number wbsle. In 'Books from Upjohn Press.', W.E. Upjohn Institute for Employment Research, 1991.

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott, "Global firms," *Journal of Economic Literature*, 2018, *56* (2), 565–619.

Bloom, Nicholas, Mirko Draca, and John Van Reenen, "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *The Review of Economic Studies*, 2016, *83* (1), 87–117.

Blundell, Richard, Rachel Griffith, and Frank Windmeijer, "Individual effects and dynamics in count data models," *Journal of econometrics*, 2002, *108* (1), 113–131.

— , — , **and John van Reenen**, "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms," *Review of Economic Studies*, 1999, *66* (3), 529–554.

Bombardini, Matilde, Bingjing Li, and Ruoying Wang, "Import Competition and Innovation: Evidence from China," Technical Report 2017.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *The Review of Economic Studies*, 06 2021. rdab030.

Bustos, Paula, "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *American Economic Review*, 2011, *101* (1), 304–340.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American economic review*, 2009, *99* (4), 1145–77.

Coelli, Federica, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "Better, Faster, Stronger: Global Innovation and Trade Liberalization," Working Paper 22647, National Bureau of Economic Research September 2016.

de Rassenfosse, Gaetan, Helene Dernis, Dominique Guellec, Lucio Picci, and Bruno van Pottelsberghe de la Potterie, "The worldwide count of priority patents: A new indicator of inventive activity," *Research Policy*, 2013, *42* (3), 720–737.

Dhingra, Swati, "Trading Away Wide Brands for Cheap Brands," *American Economic Review*, October 2013, *103* (6), 2554–2584.

Gaulier, Guillaume and Soledad Zignago, "BACI: International Trade Database at the Product-Level. The 1994-2007 Version," Working Papers 2010-23, CEPII research center October 2010.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," Working Paper 24408, National Bureau of Economic Research March 2018.

Grossman, Gene M. and Elhanan Helpman, *Innovation and growth in the global economy*, MIT Press, 1991.

— **and** — , "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 1991, *58* (1), 43.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, 2005, *36* (1), 16–38.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang, "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 2014, *104* (6), 1597–1629.

Iacovone, Leonardo, Wolfgang Keller, and Ferdinand Rauch, "Innovation Responses to Import Competition," 2011. Unpublished manuscript.

Impullitti, Giammario and Omar Licandro, "Trade, firm selection, and innovation: the competition channel," *The Economic Journal*, February 2018, *128*, 189–229.

Insee, "NAF REV. 2 et CPF REV. 2 : Guide d'utilisation," 2016.

Lequien, Matthieu, Martin Mugnier, Loriane Py, and Paul Trichelair, "Linking patents to firms: insights with French firms," 2019. Mimeo Banque de France.

Lileeva, Alla and Daniel Trefler, "Improved access to foreign markets raises plant-level productivity... for some plants," *The Quarterly journal of economics*, 2010, *125* (3), 1051–1099.

Malgouyres, Clément, Thierry Mayer, and Clément Mazet-Sonilhac, "Technology-induced Trade Shocks? Evidence from Broadband Expansion in France," CEPR Discussion Papers 13847, C.E.P.R. Discussion Papers 2019.

Mayer, Thierry and Gianmarco I. P. Ottaviano, "The Happy Few: The Internationalisation of European Firms," *Intereconomics - Review of European Economic Policy*, 2008, *43* (3), 135–148.

_ , Marc J. Melitz, and Gianmarco I. P. Ottaviano, "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, 2014, *104* (2), 495–536.

_ , _ , and _ , "Product Mix and Firm Productivity Responses to Trade Competition," NBER Working Papers 22433, National Bureau of Economic Research, Inc July 2016.

Melitz, Marc J., "Competitive effects of trade: theory and measurement," *Review of World Economics*, February 2018, *154* (1), 1–13.

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli, "Disambiguation of patent inventors and assignees using high-resolution geolocation data," *Nature Scientific Data*, 2017, *4*.

APPENDIX

2.A Data description

2.A.1 Patent data

Our first database is `PATSTAT Spring 2016` which contains detailed information about patent applications from every patent office in the world. Each patent can be exactly dated to the day of application, which is sometimes referred to as the "filing date".

Counting patent applications Each French firm is associated with a number of patent applications by that firm each year (see section 2.A.4). If the firm shares a patent with some other firms, then we only allocate a corresponding share of this patent to the firm. This raises the well-documented issue of truncation bias (Hall et al. (2005)). Indeed as we come closer to the end of the sample, we observe a smaller fraction of all patents since many of them are not yet granted.[33] In addition, there is a legal obligation to wait 18 months before publication in `PATSTAT`. With our version of Spring 2016 this implies that we can assume the data to be reasonably complete up to 2012. The sector-time fixed effects also deal with the truncation bias in our regressions. An alternative solution could be to use the year of granting instead of the year of application. However, the former is less relevant than the latter as it is affected by administrative concerns and also by potential lobbying activities that have little to do with the innovation itself. In order to be as close to the time of the innovation as possible, we follow the literature and consider the filing date. We consider every patent owned by a French firm, regardless of the patent office that granted the patent rights. Here we need to be aware of the differences in regulations across intellectual property offices. Some patent offices, especially those of Japan and Korea, offer less breadth to a patent, which implies that more patents are needed to protect a given invention than in other patent offices (see de Rassenfosse et al., 2013). Since we only consider French firms, this would become an issue only if some French firms patent relatively more in countries like Japan or Korea, which would induce an upward bias in the number of patents held by those firms. However, we use a count of priority patent applications only, which is immune to this potential bias.

Priority patent applications The fact that an inventor might want to patent its invention in different countries (or through supranational patent offices like PCT or EPO) makes it impossible to consider that one patent is equal to one invention. For this reason, patents are associated with a family which

[33] The time between patent application and patent granting is a little more than 2 years on average but the distribution of this lag is very skewed with few patent applications still waiting for patent granting many years after the application.

gather different patents which are more or less related to the same invention. More precisely, during a 12-month period following the filing of an application, the applicant has the *right of priority*. During this period, the applicant can file a similar patent in a different patent office and *claim the priority* of the first application when filing this subsequent application. If the priority claim is valid, the date of filing of the first application is considered to be the effective date of filing for the subsequent applications. This first application corresponds to the priority patent. All subsequent filings of the same intellectual property (in particular if they are in other countries) are secondary filings.

Citations We also use `PATSTAT` information on citations received by patents owned by French firms. Citations are often used to address the problem that all patents are not of equal quality and that simply counting the number of patent applications provides a noisy measure of the true innovation performance of a firm. However, the truncation bias issue is even worse with citations than with patent count. Patents from say 2010 have less time to be cited than patents from 1980 regardless of their respective qualities. Comparing different cohorts of patents can thus lead to misinterpreting what is reflected by the total number of citations received by a firm. To address this problem, we consider the number of citations received within a certain time window after the application date (usually 3 or 5 years). Using sector times year fixed effects in the regressions also helps to alleviate this concern.

2.A.2 Firm-level accounting data

Our second data source, provided by the DGFiP-Insee and called `FICUS` and `FARE`, provides us with accounting data for French firms. The corresponding data are drawn from compulsory reporting of firms and income statements to fiscal authorities in France. Since every firm needs to report every year to the tax authorities, the coverage of the data is all French firms from 1994 to 2012 with no limiting threshold in terms of firm size or sales. This dataset provides us with information on the turnover, employment, value-added, the four-digit sector the firm belongs to ... This corresponds to around 47 million observations and the number of observations per year increases from 1.9m to 3.9m over the period we consider.

The manufacturing sector is defined as category C of the first level of the NAF (*Nomenclature d'Activités Françise*), the first two digits of which are common to both NACE (Statistical Classification of Economic Activities in the European Community) and ISIC (International Standard Industrial Classification of All Economic Activities). Insee provides each firm with a detailed principal activity code (APE) with a top-down approach: it identifies the 1-digit section with the largest value added. Among this section, it identifies the 2-digit division with the largest value-added share, and so on until the most detailed 5-digit APE code (Insee, 2016). It is therefore possible that another 5-digit code shows a larger value-added share than the APE identified, but one can be sure that the manufacturing firms identified produce a larger value-added in the manufacturing section than in any other 1-digit section, which is precisely what we rely on to select the sample of most of our regressions. The 2-digit

NAF sector, which we rely intensively on for our fixed effects, then represents the most important activity among the main section of the firm. Employment each year is measured on average within the year and may therefore be a non-integer number.

A unique 9-digit identifier called *Siren number* is associated to each firm, this number is given to the firm until it disappears and cannot be assigned to another firm in the future. When a firm merges with another firm, or is acquired by another firm, or makes significant changes in its organization, this number may change over time. Hence, new entrant *Sirens* in our database do not necessary correspond to new firms.

2.A.3 Trade data

Customs data for French firms Detailed data on French exports by product and country of destination for each French firm are provided by the French Customs. These are the same data as in Mayer et al. (2014) but extended to the whole 1994-2012 period. Every firm must report its exports by destination country and by very detailed product (at a level finer than HS6). However administrative simplifications for intra-EU trade have been implemented since the Single Market, so that when a firm annually exports inside the EU less than a given threshold, these intra-EU flows are not reported and therefore not in our dataset. The threshold stood at 250 000 francs in 1993, and has been periodically reevaluated (650 000 francs in 2001, 100 000 euros in 2002, 150 000 euros in 2006, 460 000 euros in 2011). Furthermore flows outside the EU both lower than 1 000 euros in value and 1 000 kg in weight are also excluded until 2009, but this exclusion was deleted in 2010.

Country-product bilateral trade flows CEPII's database `BACI`, based on the UN database `COMTRADE`, provides bilateral trade flows in value and quantity for each pair of countries from 1995 to 2015 at the HS6 product level, which covers more than 5,000 products.

2.A.4 Matching

Our paper is the first to merge those three very large - patent, administrative, and customs - datasets covering exporting French firms. Merging administrative firm-level data from FICUS/FARE and Customs data is fairly straightforward as a firm can be identified by its *Siren* identifier in both datasets.[34] Thus the main challenge is to match either of these two datasets with `PATSTAT`. Indeed, `PATSTAT` only reports the name of the patent applicant(s). Not only can this name be slightly different from the name reported in the other two databases, but it may also change over time, for example because

[34] Although one must keep track of the different definitions of firms across these two datasets.

of spelling mistakes. We thus relied on the work of Lequien et al. (2019) who developed a matching algorithm to map patents with the corresponding French firms.

Lequien et al. (2019) proceed in three main steps to merge PATSTAT and SIRENE:

1. For each *Siren* number from SIRENE, find a small subset of applicant firms in Patstat with phonetic similarities:
    - perform cleaning, splitting and phonetic encoding on firms' name in both databases. Too common words are deleted (THE, AND, CO, FRANCAISE . . .).
    - sort each name by least frequent encoding in SIRENE. The more often a word appears in the database, the less information it can convey to identify firms.
    - for each SIRENE firm, the first (ie least frequent) cleaned word of the firm's name is compared with every PATSTAT name. All the PATSTAT names containing this word form a first subset of possible matches. Then the second word of the firm's name is compared with every name in this subset, reducing it further. This procedure stops before arriving at a null subset, and yields a set of likely PATSTAT matches for each SIRENE name. Very often this set is null because the majority of firms do not patent. On average, this subset contains 10 applicants, reducing a lot the computationally intensive comparisons.
2. Computation of parameters on these possible matches
    - Comparison of the names (raw names, and cleaned names), using Levenshtein distances and an inclusion parameter (all the words in one name are included in the name from the other database)
    - zip code comparison (*code postal*)
    - date comparisons (a firm cannot have patented before its creation)
3. Matching with supervised learning
    - Sample from INPI (*Institut National de la Propriété Intellectuelle*) with 15,000 true matches between *Siren* number and PATSTAT *person id* (and in total 170,000 pairs, with the corresponding known mismatches).
    - This sample is randomly split into a learning sample and a verification sample (this procedure is repeated 10 times, and the recall and precision measures are averaged over them, so that the choice of the sample does not alter the results). This allows to choose the relevant variables and estimate the parameters.
    - apply this model on all the possible matches identified in the previous step.
    - in 90% of cases, unique matching. In the remaining 10% of cases, filter further with a decision tree (is the date of creation of the firm lower than the first filing of the applicant?,

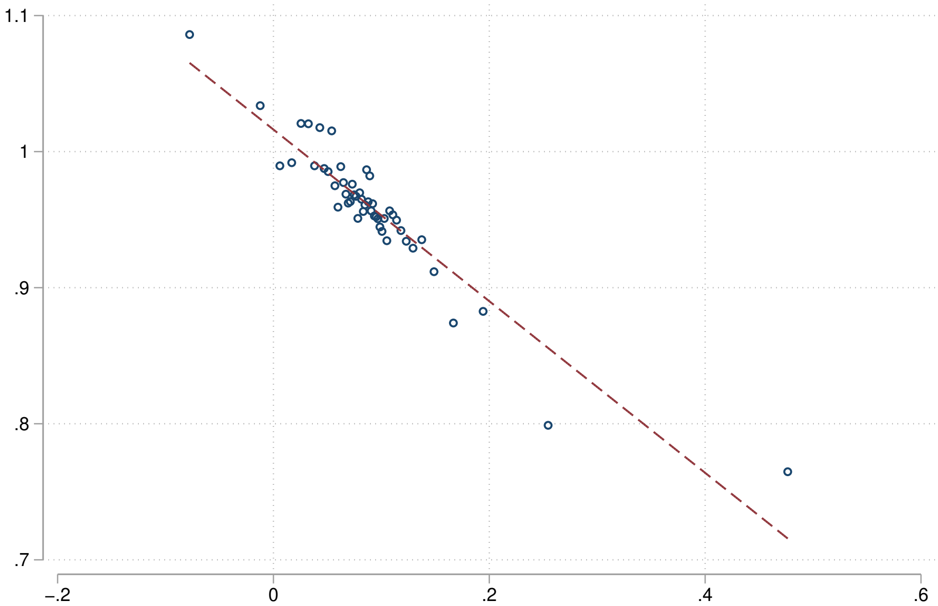
which couple has the minimum Levenshtein distance between raw names, between cleaned names, is one of the names included in the other?, which firm has the maximum number of employees?)

Based on the (rotating) verification sample taken from INPI data, the recall rate (share of all the true matchings that are accurate) is at 86.1% and the precision rate (share of the identified matches that are accurate) is at 97.0%.

2.B Time lag in exports reporting between production and customs data

The different timing for recording the export transaction between tax and customs authorities materializes in the annual data in particular when the transaction occurs at the end of a year t – it is recorded in the tax data for year t – but the shipment occurs at the beginning of the following year, in which case it is recorded in the customs data in year $t+1$. Because part of the January $t+1$ (customs) exports is recorded as year t (tax) exports, a firm with larger (customs) exports in January of year $t+1$ is expected to show a larger discrepancy between tax and customs exports in year t. Figure 2.13 reports the bin-scatter of the ratio of customs over production exports in year t (y axis) versus the share of January $t+1$ exports over exports in year t (both from the customs data, x axis), absorbing firm fixed effects. It shows that when January $t+1$ (customs) exports represent a bigger share of year t (customs) exports, then the customs data falls shorter than the production data for year t.

Figure 2.13: CUSTOMS/PRODUCTION DISCREPANCY IN YEAR t VERSUS $t+1$ JANUARY SHARE OF YEAR t EXPORTS



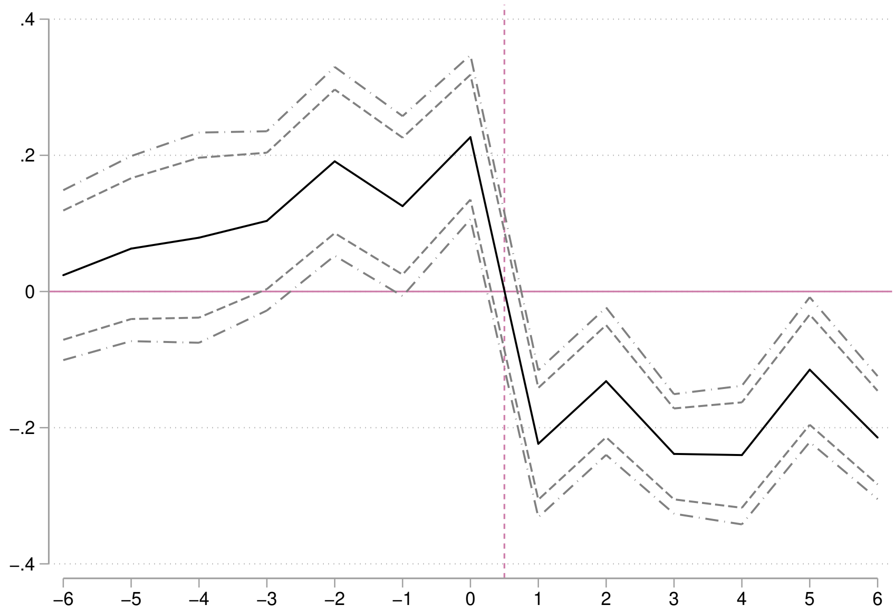
Notes: This Figure reports the bin-scatter of the ratio of customs over production exports for year t (y axis) against the ratio of January $t+1$ exports over year t exports (both taken from customs data). Firm fixed effects are absorbed. Number of observations: 53,287. Years: 1994-2012

We extend this analysis over the last months of year t and the first months of year $t+1$ with the following regression:

$$X_{f,t}/X^*_{f,t} = \sum_{m=-6}^{6} \alpha_m \frac{X_{f,m}}{X_{f,t}} + \mu_{s(f,t),t} + \nu_f + \varepsilon_{f,t} \tag{2.9}$$

where $X_{f,m}$ is (customs) exports for month m of year $t+1$ if $m > 0$, or month $12+m$ of year t if $m \leq 0$. 0 corresponds to December t, 1 to January $t+1$. We control for firm fixed effects and the sector of the firm. We keep in the regression only observations where $X_{f,t}/X^*_{f,t} \leq 10$ and where each share $\frac{X_{f,m}}{X_{f,t}} \leq 1$. Figure 2.14 reports the coefficients α_m along with their 95 and 99 confidence intervals (standard errors are clustered at the firm level). Everything else equal, if the first months of year $t+1$ represent a larger share of year t exports, then the ratio of yearly exports from customs to production data is smaller. Conversely if the last months of year t represent a larger share of yearly exports, then the customs yearly figure is bigger relative to the production figure.

Figure 2.14: DIFFERENCE IN REPORTING TIMING BETWEEN CUSTOMS AND PRODUCTION SOURCES

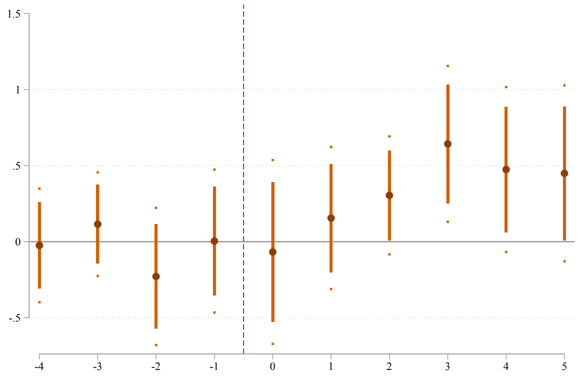


Notes: This Figure reports the coefficients α_m and corresponding 95% and 99% confidence intervals from equation (2.9). Number of observations: 58,027. Years: 1994-2012

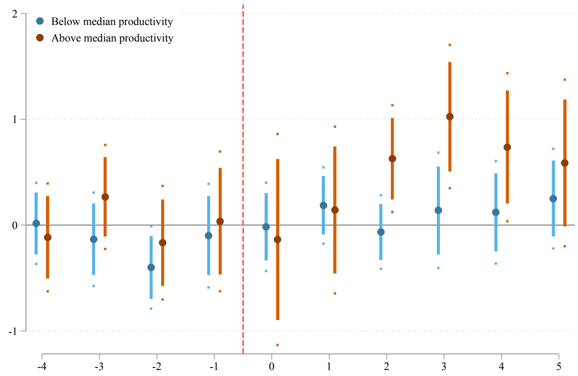
2.C Additional Empirical results

Figure 2.15: OLS: OTHER VARIABLES

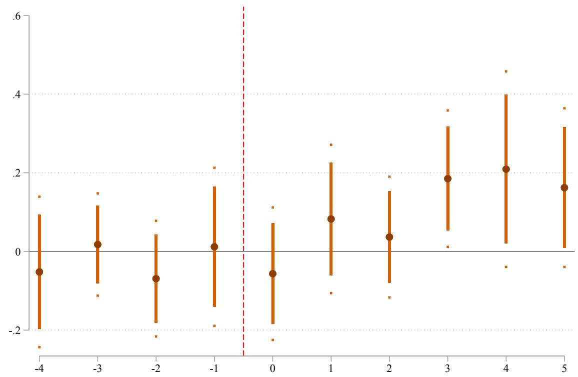
(a) Citations within 3 years



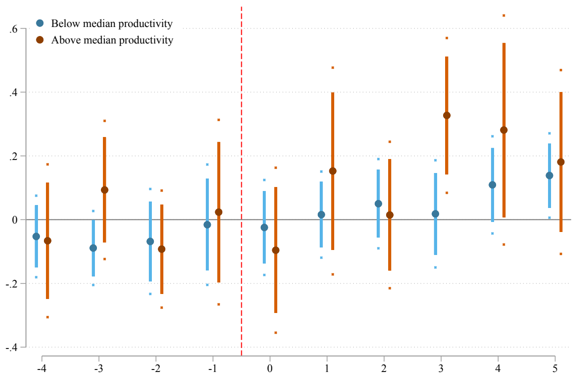
(b) Citations within 3 years



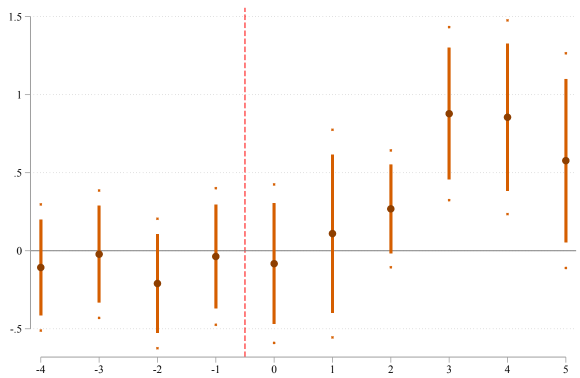
(c) Top 10% patents



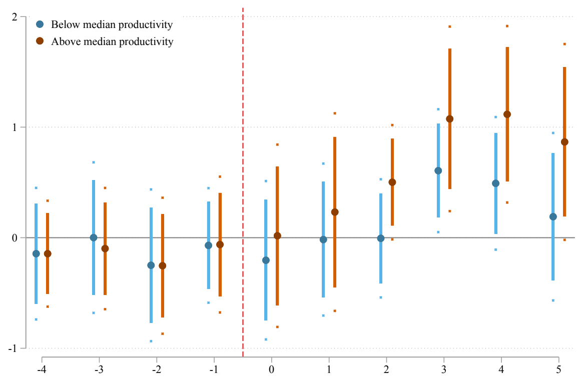
(d) Top 10% patents



(e) All patents



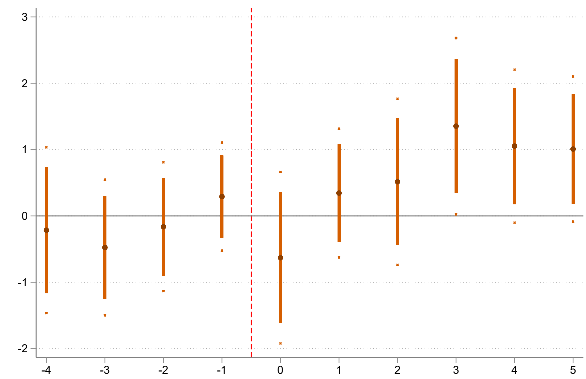
(f) All patents



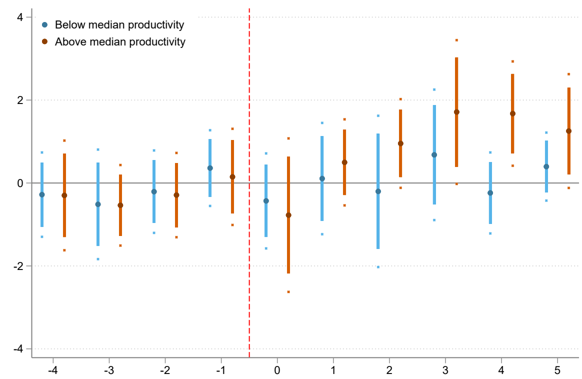
Notes: These Figures replicate Figures 2.4a and 2.6a but using different measures of innovation as the dependent variable: respectively counting citations received within a 3 year window, counting the number of patents among the 10% most cited in the year and counting any patent (whether priority or secondary filing). Number of observations: 22,175.

Figure 2.16: INVERSE HYPERBOLIC SINE TRANSFORMATION OF THE DEPENDENT VARIABLE

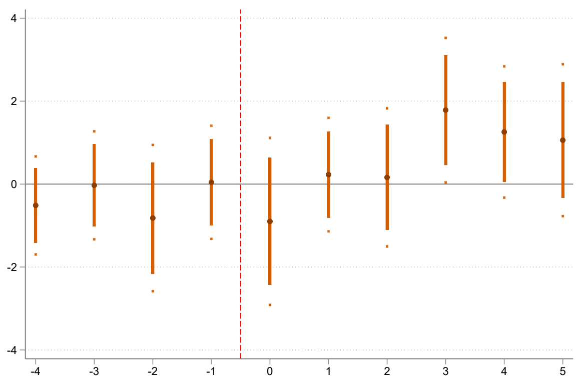
(a) Priority patents



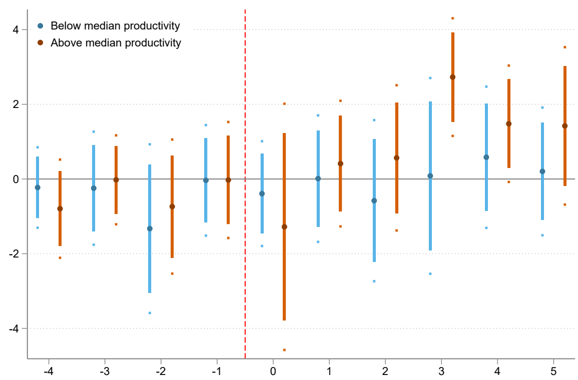
(b) Priority patents



(c) Citations within 5 years



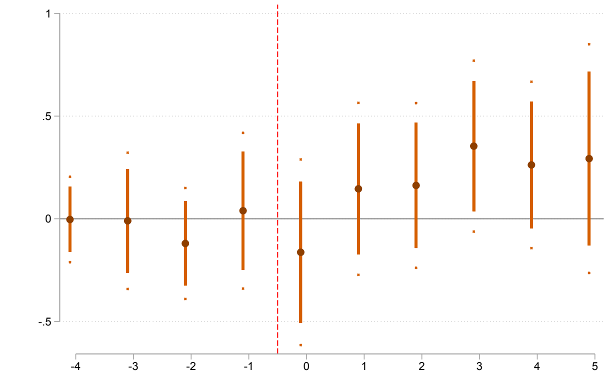
(d) Citations within 5 years



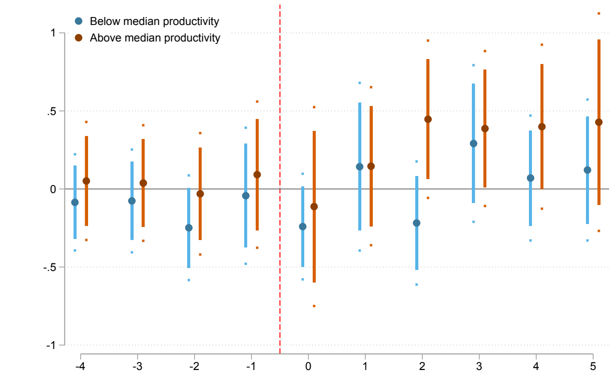
Notes: These Figures replicate Figures 2.4 and 2.6 but use the inverse hyperbolic sine transformation (asinh) of the number of priority patents or of 5-year citations as the dependent variable. Number of observations: 22,175

Figure 2.17: OLS: REMOVING THE MARKETS WHERE A FRENCH FIRM IS A LEADER

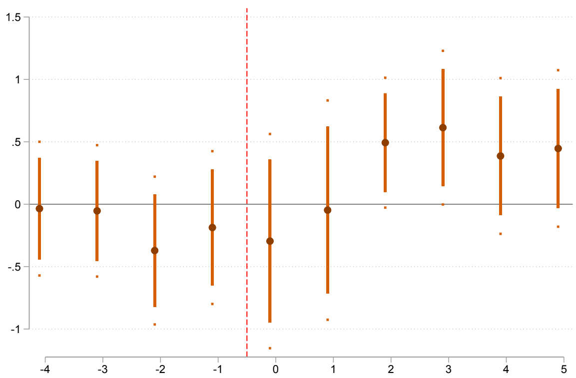
(a) Priority patents



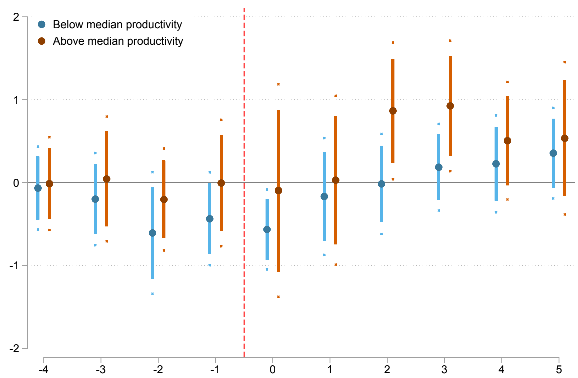
(b) Priority patents



(c) Citations within 5 years

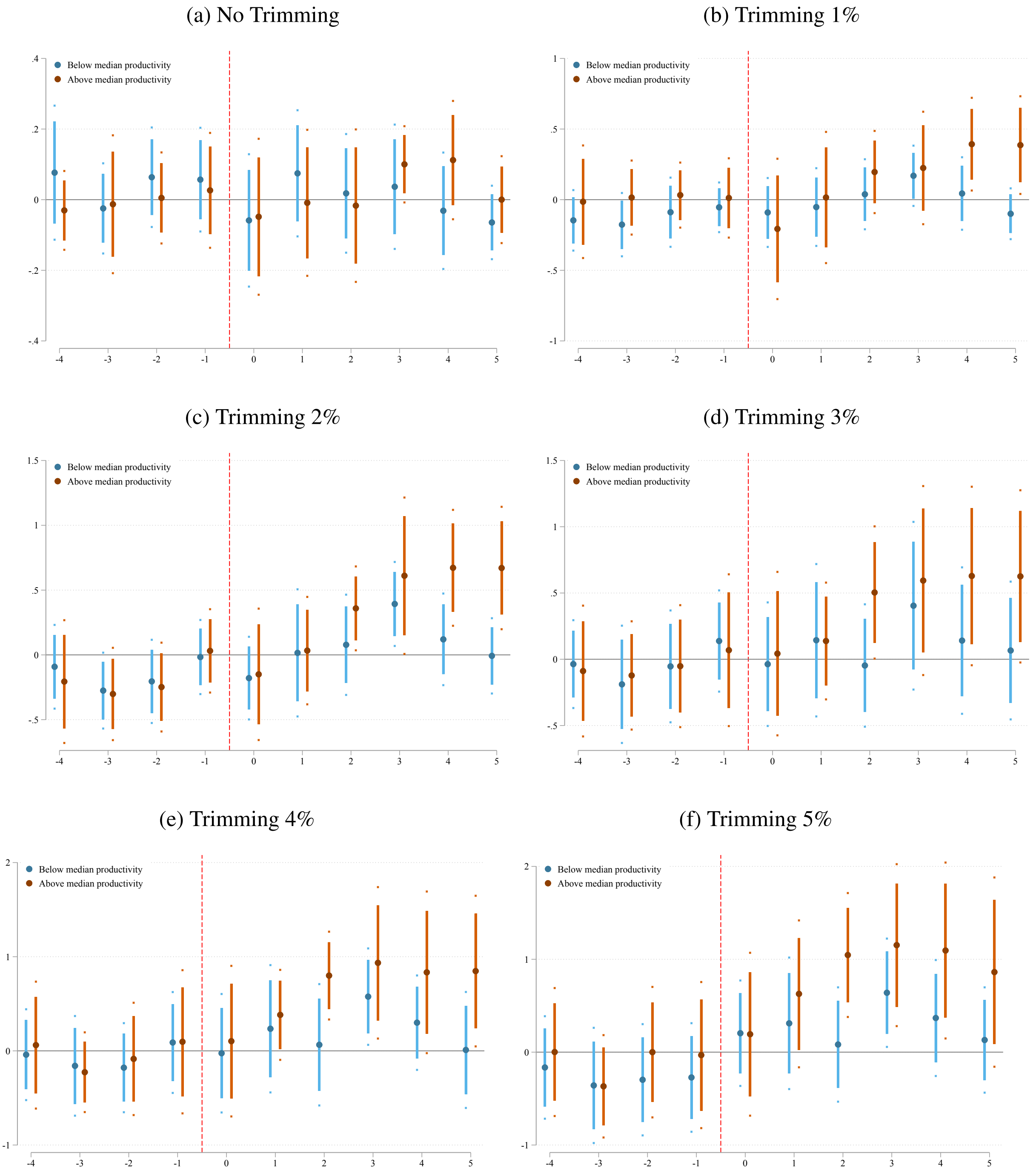


(d) Citations within 5 years



Notes: These Figures replicate Figures 2.4 and 2.6 but dropping firm-destination pairs whenever the firm's market share in the destination exceeds 10% in the construction of the demand shock variable. Number of observations: 21,859

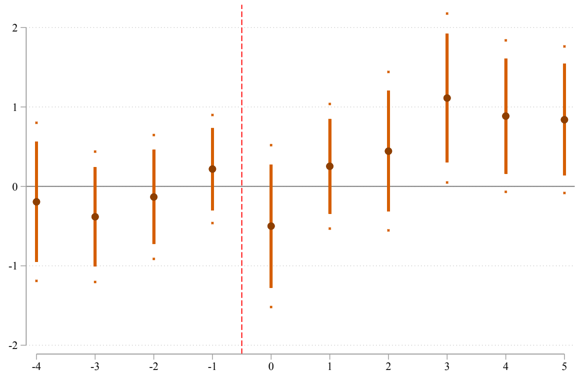
Figure 2.18: Different trimming thresholds - OLS: Priority patents



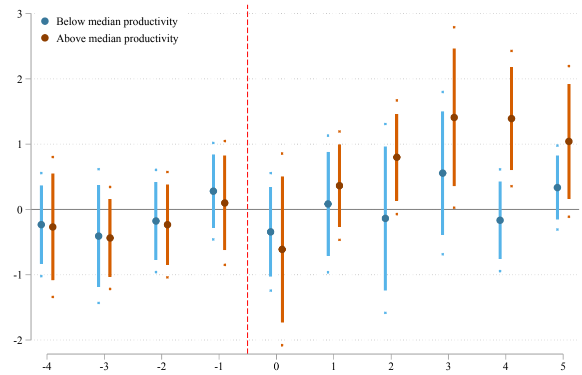
Notes: These Figures replicate Figure 2.6a but use a different trimming of the demand shocks, respectively: 0, 1, 2, 3, 4 and 5%. Number of observations: 26,954; 24,866; 23,087; 21,381; 19,828 and 18,475 respectively.

Figure 2.19: OLS: SAMPLE OF FIRMS THAT INNOVATED BEFORE 1994

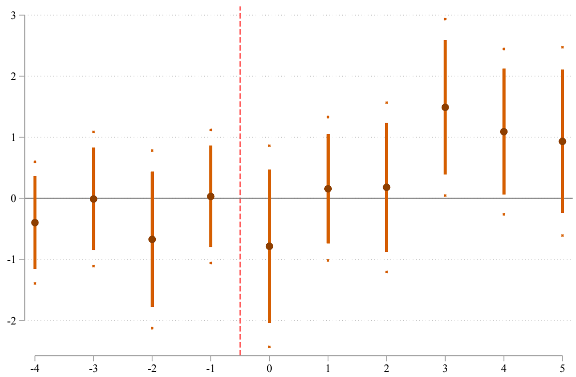
(a) Priority patents



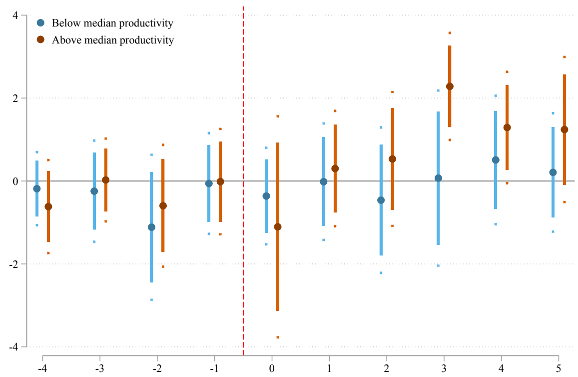
(b) Priority patents



(c) Citations within 5 years



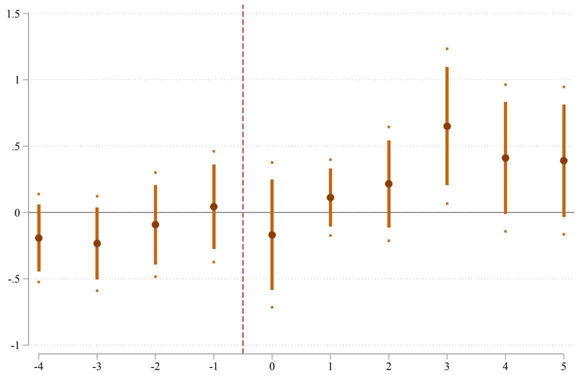
(d) Citations within 5 years



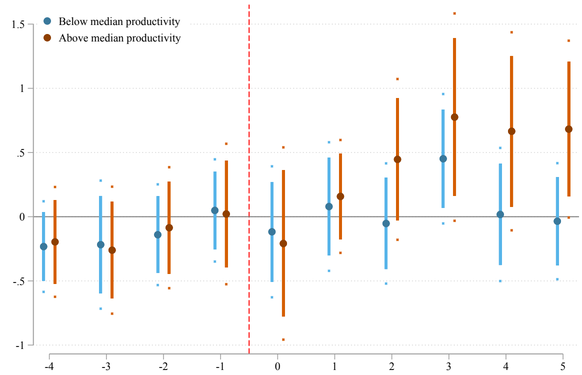
Notes: These Figures replicate Figures 2.4 and 2.6 but restricting to firms that innovated before 1994 (i.e. with a patent application with filing date between 1985 and 1994). Number of observations: 6,866

Figure 2.20: OLS: SAMPLE OF FIRMS WITH $t_0 = 1994$

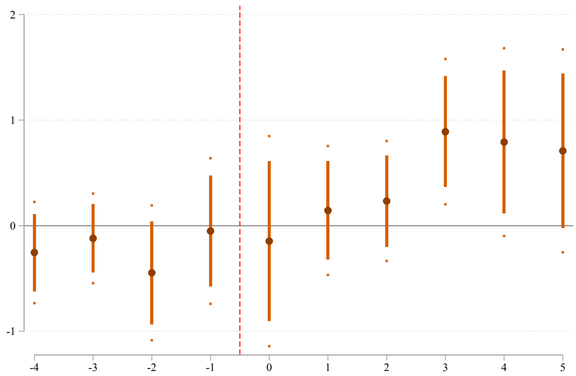
(a) Priority patents



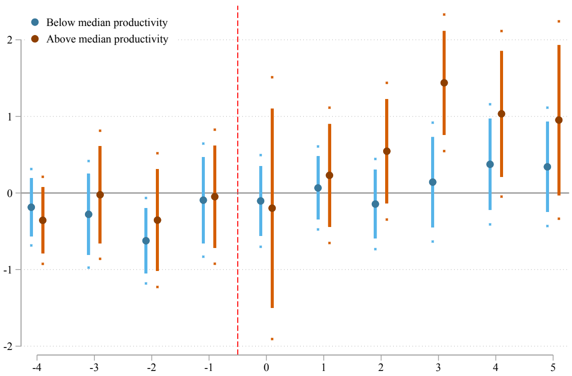
(b) Priority patents



(c) Citations within 5 years



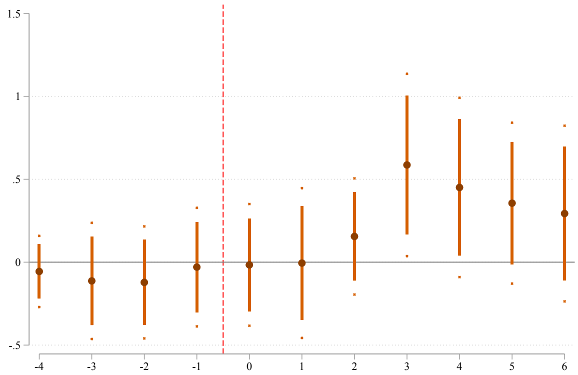
(d) Citations within 5 years



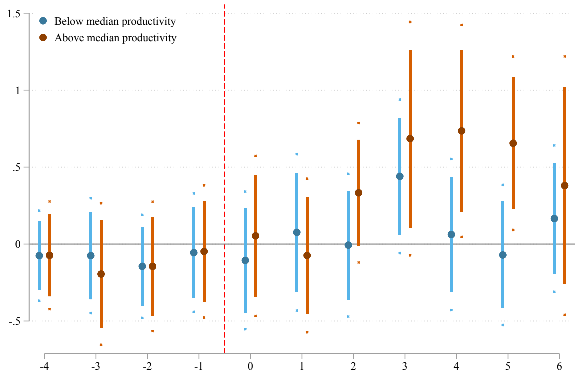
Notes: These Figures replicate Figures 2.4 and 2.6 but restricting to firms that first exported before 1994 ($t0 = 1994$ in equations (2.2) and (2.4)). Number of observations: 15,742

Figure 2.21: OLS: 6 LAGS

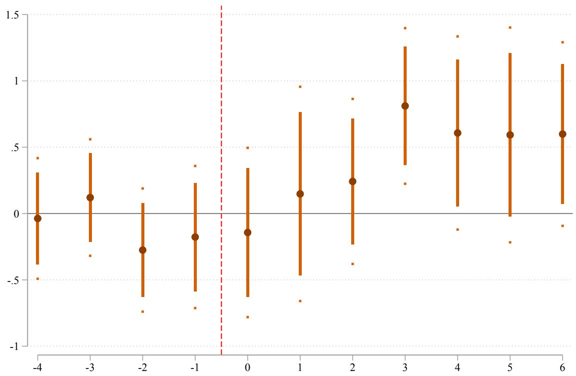
(a) Priority Patents



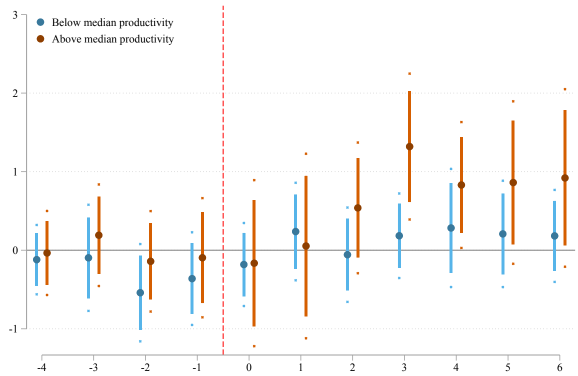
(b) Priority Patents



(c) Citations within 5 years



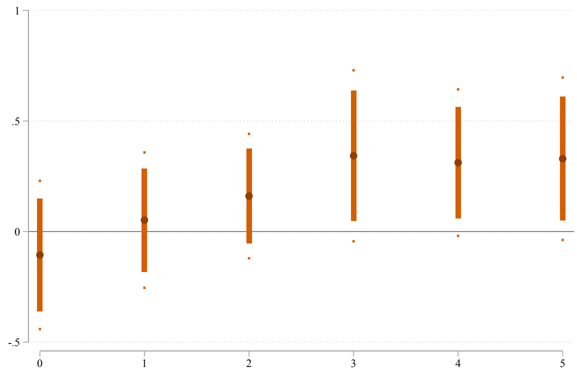
(d) Citations within 5 years



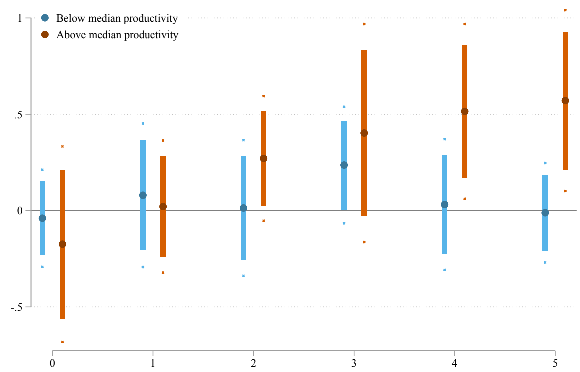
Notes: These Figures replicate Figures 2.4 and 2.6 but using 6 lags and 4 leads (therefore defining $k = 6$ and $k' = 4$ in equations (2.2) and (2.4)). Number of observations: 18,707

Figure 2.22: OLS: NO PRE-TREND

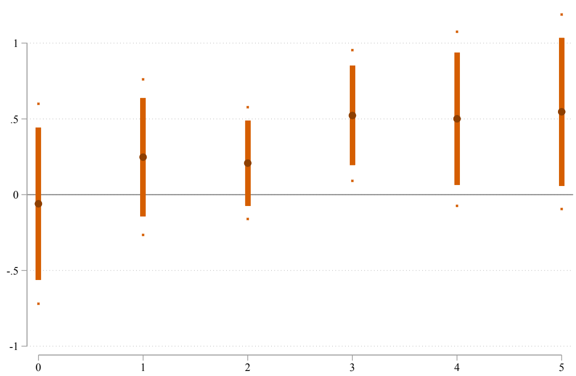
(a) Priority Patents



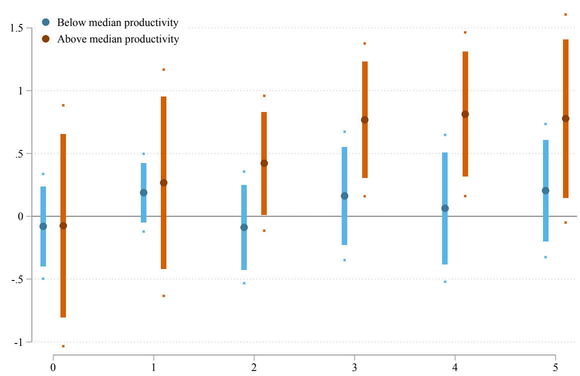
(b) Priority Patents



(c) Citations within 5 years



(d) Citations within 5 years



Notes: These Figures replicate Figures 2.4 and 2.6 but without using any lead (therefore defining $k = 5$ and $k' = 0$ in equations (2.2) and (2.4)). Number of observations: 25,237

Table 2.8: DETAILED REGRESSION RESULTS

	One group		Two groups	
	OLS	NB	OLS	NB
Shock $D_{f,t}$				
at t+4	-0.059 (0.100)	0.321 (0.874)		
Below Median			-0.062 (0.116)	0.602 (0.957)
Above Median			-0.097 (0.150)	-0.005 (0.762)
at t+3	-0.140 (0.119)	-0.818 (0.524)		
Below Median			-0.139 (0.156)	-0.729 (0.802)
Above Median			-0.174 (0.153)	-0.935 (0.655)
at t+2	-0.091 (0.126)	-0.241 (0.804)		
Below Median			-0.142 (0.122)	-0.028 (0.855)
Above Median			-0.101 (0.138)	-0.596 (0.726)
at t+1	0.901 (0.136)	-0.061 (0.795)		
Below Median			0.097 (0.126)	0.813 (0.918)
Above Median			0.049 (0.177)	-0.538 (1.081)
Total pre-trend	-0.199 (0.328)	-0.799 (1.588)		
Below Median			-0.247 (0.328)	0.658 (2.016)
Above Median			-0.323 (0.381)	-2.074 (1.619)
at t	-0.072 (0.151)	-1.172 (1.049)		
Below Median			-0.101 (0.144)	-1.481 (1.178)
Above Median			-0.449 (0.218)	-0.764 (1.264)
at t-1	0.118 (0.143)	0.143 (0.746)		
Below Median			0.105 (0.195)	1.189 (1.211)
Above Median			0.135 (0.151)	-0.403 (0.777)
Total current	0.046 (0.257)	-1.029 (1.580)		
Below Median			0.004 (0.251)	-0.292 (1.401)
Above Median			0.904 (0.300)	-1.167 (1.824)
at t-2	0.240* (0.128)	1.385** (0.666)		
Below Median			0.017 (0.147)	0.119 (1.253)
Above Median			0.439** (0.178)	2.393*** (0.733)
at t-3	0.573*** (0.192)	2.301** (1.107)		
Below Median			0.406 (0.162)	1.485 (1.621)
Above Median			0.687** (0.265)	2.393*** (0.734)
at t-4	0.457** (0.163)	1.829** (0.907)		
Below Median			0.129 (0.187)	0.615 (1.511)
Above Median			0.697*** (0.243)	2.485** (1.144)
at t-5	0.392** (0.162)	1.487 (1.148)		
Below Median			0.0163 (0.171)	-0.458 (1.598)
Above Median			0.668*** (0.204)	2.539** (1.144)
Total Future	1.662*** (0.534)	7.002** (3.112)		
Below Median			0.569 (0.353)	1.761 (3.046)
Above Median			2.491*** (0.713)	10.223*** (3.004)
Initial Stock of patent (log)	0.755*** (0.0793)	1.301*** (0.110)	0.747*** (0.067)	1.277 (0.124)
Initial Employment (log)	-0.005 (0.018)	-0.049 (0.100)	0.017 (0.019)	0.291 (0.104)
Initial Sales (log)	0.050 (0.016)	0.429*** (0.064)	0.028* (0.015)	0.347 (0.070)
Initial Employment variation	0.049 (0.037)	0.234 (0.209)	0.044 (0.038)	0.201 (0.212)
Initial Sales variation	0.040** (0.018)	0.350*** (0.112)	0.042*** (0.015)	0.366 (0.095)
Fixed Effects				
Sector × year	✓	✓	✓	✓
Productivity group × year			✓	✓
Initial export intensity × year	✓	✓	✓	✓
Obs.	22,175	22,175	22,175	22,175
R2	0.292		0.294	

Notes: This table reports point estimate and standard errors (under parentheses) of all coefficients of equations (2.2) (col 1), (2.3) (col 2), (2.4) (col 3) and (2.5) (col 4). Firms are either taken all together (columns 1 and 2) or grouped according to their initial level of productivity (Above Median and Below Median, columns 3 and 4). In addition to the estimation of each coefficients, the Table also reports the Wald test on different linear combinations of coefficients (see Tables 2.6 and 2.7). ***, ** and * indicate p-value below 0.01, 0.05 and 0.1 respectively. Standard errors are clustered by groups of sector (2 digit) and group of productivity.

2.D Theoretical appendix

We describe how the equilibrium competition level λ in destination D is endogenously determined and show that λ increases with L. Although this equilibrium involves all the firms operating in D, including both the French exporters to D along with the domestic producers in D, we show that the equilibrium competition level λ is determined independently of the export supply to D (which then only impacts the number of domestic entrants and producers).

Let $\Gamma_D(\tilde{c})$ denote the cumulative distribution of baseline costs $\tilde{c}$ among domestic producers in D. We assume that $\Gamma_D(\tilde{c})$ has support on $[\tilde{c}_{0D}, +\infty)$ with $\tilde{c}_{0D} > \tilde{c}_{\min}$. Let F_D denote the fixed production cost faced by those domestic firms in D. Since a firm's operating profit is monotonic in its baseline cost $\tilde{c}$, producing for the domestic market D is profitable only for domestic firms with a baseline cost $\tilde{c}$ below a cutoff value $\widehat{C}_D$ defined by the zero profit condition:

$$\Pi(\widehat{C}_D, 0; \lambda) = F_D, \tag{ZCP}$$

where we have assumed that $\widehat{C}_D > \widehat{C}_I$ so that the firm with the cutoff cost $\widehat{C}_D$ does not innovate (and hence does not incur any innovation cost). Entry is unrestricted subject to a sunk entry cost F_D^E. In equilibrium, the expected profit of a prospective entrant will be equalized with this cost, yielding the free-entry condition:

$$\int_{\tilde{c}_{0D}}^{\widehat{C}_D} [\Pi(\tilde{c}, k(\tilde{c}; \lambda); \lambda) - F_D] \, d\Gamma_D(\tilde{c}) = F_D^E. \tag{FE}$$

Proposition 1 *The two conditions* (ZCP) *and* (FE) *jointly determine a unique pair* $(\lambda, \widehat{C}_D)$.

PROOF *Uniqueness*: in $(\widehat{C}_D, \lambda)$ space, the (ZCP) condition is strictly downward-sloping while the (FE) condition is strictly upward-sloping, ensuring uniqueness of the equilibrium if such an equilibrium exists. More precisely: (a) an increase in competition from λ to $\lambda + d\lambda$ reduces the profit of firms with baseline cost $\widehat{C}_D(\lambda)$, so that those firms no longer operate; this means that $\widehat{C}_D(\lambda + d\lambda) < \widehat{C}_D(\lambda)$, which proves that the (ZCP) curve is strictly downward-sloping; (b) an increase in competition from λ to $\lambda + d\lambda$ reduces the profit of all firms (the envelope theorem ensures that at the optimal innovation level $\frac{\partial \Pi}{\partial k} = 0$ so that $\frac{d\Pi}{d\lambda} = \frac{\partial \Pi}{\partial \lambda} < 0$); this in turn means that $\widehat{C}_D$ has to strictly increase for the (FE) condition to hold, which proves that the (FE) curve is strictly upward-sloping.

Existence: We show that the (FE) curve lies below the (ZCP) curve for values of $\widehat{C}_D$ close to $\tilde{c}_{0D}$, and that the (FE) curve ends up above the (ZCP) curve for high values of $\widehat{C}_D$. As $\widehat{C}_D$ becomes close to $\tilde{c}_{0D}$, (ZCP) implies a value for λ which is positive and bounded away from zero, whereas (FE) requires λ to become arbitrarily small, because the integrand must go to $+\infty$ for the integral over a very small interval to remain equal to F_D^E. Next, recall that the (ZCP) curve must remain below the $\lambda = \frac{\alpha}{\widehat{C}_D}$ curve. Given that $\frac{\alpha}{\widehat{C}_D} \longrightarrow 0$ when $\widehat{C}_D \longrightarrow +\infty$, the $\frac{\alpha}{\widehat{C}_D}$ curve must cross the (FE) curve at some point. At this point, the (ZCP) curve lies below the (FE) curve.

For simplicity, we have abstracted from any export profits for the domestic firms. This is inconsequential for our prediction that the equilibrium competition level λ increases with market size L, so long as destination D is small relative to the size of the global export market.[35]

Proposition 2 *An increase in market size L in D leads to an increase in competition λ.*

PROOF We prove this proposition by contradiction. If λ were to decrease, then the cutoff $\widehat{C}_D$ would have to increase (see (ZCP)). Since $\pi(c;\lambda)$ is decreasing in λ, then $\Pi(\tilde{c},k;\lambda)$ must also increase for any given innovation level k when λ decreases. Given the optimization principle, $\Pi(\tilde{c},k(\tilde{c};\lambda);\lambda)$ must also increase. This, together with an increase in the cutoff $\widehat{C}_D$, represents a violation of the (FE) condition. Thus competition λ must increase when L increases.

[35]More precisely, the free entry condition can be extended to incorporate the (net) export profits Π_{-D} earned in other destinations:

$$\int_{\tilde{c}_{0D}}^{\widehat{C}_D} \left[\Pi(\tilde{c},k(\tilde{c};\lambda);\lambda) - F_D + \Pi_{-D}(\tilde{c},k;\{\lambda_{-D}\})\right] d\Gamma_D(\tilde{c}) = F_D^E,$$

where $\{\lambda_{-D}\}$ denotes the vector of competition levels in countries other than D. So long as these competition levels $\{\lambda_{-D}\}$ do not respond to changes in D, the export profits shift up the marginal benefit of innovation in (FOC) by an amount that does not depend on λ or L. This marginal benefit curve will remain an increasing function of innovation k and will shift up with any market-wide change in D that increases firm output $Q(\tilde{c},k;\lambda)$ at fixed innovation k.

Chapter 3

Tax Evasion and The Quest for Simplicity: Evidence from the Self-Employed in France

Tax Evasion and The Quest for Simplicity: Evidence from the Self-Employed in France

PHILIPPE AGHION UFUK AKCIGIT

MATTHIEU LEQUIEN STEFANIE STANTCHEVA

Addresses - Aghion: Collège de France, LSE and PSE; Lequien: Insee, Banque de France and PSE; Akcigit: University of Chicago, NBER, and CEPR; Stantcheva: Harvard University, NBER, and CEPR. Stantcheva's work is supported by the National Science Foundation under Grant CAREER No.1654517. Aghion acknowledges support from an Idex grant from *Paris-Sciences-et-Lettres* (PSL) and by a public grant overseen by the French National Research Agency (ANR) as part of the *Investissements d'Avenir* program (reference: ANR-10-EQPX-17 - *Centre d'accès sécurisé aux données* – CASD). Matthieu Lequien and Philippe Aghion thank the support of the EUR grant ANR-17-EURE-0001. We thank Alex Gelber, Larry Katz, Henrik Kleven, Magne Mogstad, Andreas Peichl, Dina Pomeranz, Emmanuel Saez, Antoinette Schoar, Joel Slemrod, and seminar participants at Mannheim, Sciences Po, and the NBER Entrepreneurship group, for feedback and comments. We thank Maxime Gravoueille, Vlad Ciornohuz, Antoine Mayerowitz, Simon Bunel, Francois-Xavier Ladant, and Cyril Verluise for outstanding research assistance.

Abstract

We analyze how French self-employed individuals respond to a notch in their tax schedule, and we identify the motives behind that response. French self-employed can choose between three fiscal regimes, each of which differs in terms of tax incentives and administrative simplicity. In the two simplest regimes, we show a massive bunching just below the turnover threshold above which the individuals no longer qualify for those regimes. Entrepreneurs respond to tax incentives, as individuals with higher marginal tax rates bunch more. Many behavioral patterns suggest that this observed bunching comes at least to some extent from misreporting: just below the threshold, individuals report more often round numbers and appear to shift income to their partner when both are self-employed. In addition, we show that individuals learn over time how the system works – and how to take advantage of it. Finally a structural model of regime choice, estimated by fitting data moments, confirms the role of tax evasion and shows that the self-employed place a high value on tax simplicity, particularly for the super simplified regime.

JEL codes: H26, H24, H21

Keywords: Taxation, Bunching, Evasion, Complexity, Self-Employment, Entrepreneurship.

3.1 Introduction

"Simplicity is the ultimate sophistication," wrote Leonardo da Vinci. Many policy makers would probably agree with his statement. Designing a policy that fulfills its stated goals, provides clear and correct incentives without unintended consequences, minimizes administrative hassle for individuals, and at the same time remains sufficiently simple for people to understand is an enormous challenge. Tax policy is a case in point: the best tax incentives may turn out to be ineffective if people do not understand them. Even worse, complexity may make the system more regressive if it is mostly the least sophisticated agents or those who cannot afford professional tax advice who cannot understand it and benefit from it. Many tax and transfer policies are targeted towards the bottom of the income distribution, where simplicity may be even more important, and where complexity may prevent the very same people targeted by these policies from taking advantage of them.

In this paper, we try to disentangle the extent to which the choice of tax regime by self-employed individuals in France, is driven by a preference for simplicity or simply by the incentive to evade taxation. We define tax simplicity as the combination of conceptual simplicity and practical simplicity: a system is simple if it is both easy to understand and logistically easy to handle.

The self-employed are a very interesting group to study *per se*. They have become more numerous and important in recent years, through the rise of platforms such as Uber, Air BnB, or Task Rabbit, and the outsourcing of jobs previously done in-house. In recent work, Katz and Krueger (2016) and Katz and Krueger (2017) cast light on the rise of alternative work arrangements – those differing from conventional self-employment and regular employment – and on the ensuing fragmentation of the labor market.

Our own focus on the self-employed is mainly driven by the following considerations. First, they are typically shown to be much less constrained than wage earners and can more easily adjust their incomes to tax incentives (Saez, 2010; Kleven and Waseem, 2013). This is important if we want to measure how people respond to simpler or more complex tax policies. Second, since the self-employed are their own decision makers, there is a more direct map between their own understanding of the tax system and their response to it. This link is weakened for wage earners, since it may be their company determining their pay structure and responses to taxes, based on its own (presumably, better) knowledge of the tax system. Third, self-employment in France serves as a particularly well-suited quasi-laboratory for studying the effects of tax simplicity and complexity. Indeed, it has a very unique variety of fiscal "regimes"– or modes of taxation of self-employment– which differ not only in their monetary incentives, but also in their degree of tax simplicity. These fiscal regimes have changed significantly over time, offering the opportunity to study learning and dynamic adjustments. They also impact different groups of agents heterogeneously, thus providing valuable policy variation that helps our estimation.

Our first main contribution is to introduce and use new individual tax returns data from the French internal revenue service over the period 1994-2015. The tax returns data is combined with additional administrative and large-scale survey data, to yield information on employment, demographics, education and government benefits received. This highly valuable combination of administrative tax data and census-style survey data allows us to study the characteristics of agents who respond differently to tax incentives.

In Section 3.2, we start by describing the landscape of French policies related to self-employment. There are three regimes under which the self-employed may choose to operate, which differ along two main dimensions: monetary tax incentives and tax simplicity. In brief, the "standard regime" treats an individual's net business income (revenues minus costs) as taxable income, which is advantageous for businesses with employees, significant investments, or high operating costs. It does, however, come with the most involved and costly tax accounting requirements, which also limit the scope for misreporting. The "simplified regime" cuts down on tax hassle and allows agents to claim a flat-rate rebate as a fraction of revenues instead of reporting their true business costs, which can be very advantageous for agents with low operating costs. The "super simplified" regime further increases tax simplicity by replacing all income taxes and social insurance contributions by a unique – and relatively low– flat rate payment proportional to gross revenues. The simplified and super simplified regimes require that revenues are below an eligibility threshold. This threshold depends on the type of business activity, and has changed over time. Thus, broadly speaking, the simplified and super simplified regimes are well suited to agents with small and slow-growing activities, with relatively low operating costs and investments, and with strong preferences for tax simplicity.

In Section 3.3, we provide key new summary statistics on the self-employed for the period 1994-2015.[1] Section 3.4 formally models the three self-employed regimes, their financial (net-of-tax) payoffs, and the costs imposed by tax requirements, which we call "tax hassle costs." The eligibility thresholds create a special type of discontinuity, not only in monetary payoffs, but also in tax simplicity. We express this discontinuity – or "notch"– in monetary terms as a function of underlying parameters, such as an agent's tax bracket, activity type, tax hassle costs and operating costs. We find very significant behavioral responses (in terms of regime choice and income) to the notches created by the eligibility thresholds. We also highlight heterogeneity in responses: Agents who have other sizable sources of income, such as salaried income or pension income, exhibit much stronger bunching. The same holds true for agents who stand to gain more from fiscal optimization, namely those in higher tax brackets. Importantly, only agents with a at least a high school degree respond to the eligibility thresholds; those without one do not.

[1]The self-employed are on average older than wage earners, more likely to be retired, more educated, more likely to be in high skill occupations, and have higher labor, capital, and total income. They are less likely to receive unemployment or social insurance benefits. The fraction of agents with self-employed income remained stable at around 5% of all tax filers aged 18-65 until 2009 and has risen since then. The fraction of agents who earn only self-employed income remained at 4% until 2009 and has increased sharply since then.

Our second main contribution is to try and explain the bunching. We proceed in two main steps. Step one, in Section 5: there, we provide three pieces of evidence to the effect that bunching is at least partly associated with tax evasion. First piece of evidence: revenue statements are more often round numbers close to the thresholds than far from them, an indication that the reported figure is more likely to have been forged. Second piece of evidence: in households with two self-employed individuals, the highest earner appears to shift some of her income to her partner as she approaches the threshold. Third piece of evidence: bunching increases over time, and agents react rapidly to a large and salient institutional change whereas they respond slowly to a smaller or more complex change.

Step two, in Section 6: we try to disentangle between pure monetary motives and the quest for simplicity to explain individuals' attempt to file under the simpler regimes. To disentangle between the two motives, we estimate a structural model by fitting data moments. Our estimation confirms the role of tax evasion. But it also shows that tax simplicity is at work in explaining the observed bunching. Moreover, the simplicity motive appears to be particularly important in the super simplified regime, where it is valued at around 500 € (versus less than 100 € in the simplified regime).

Our paper relates to several strands of literature. First, the literature on the effects of taxation on entrepreneurship and self-employment. Cullen and Gordon (2007) use U.S. tax returns data to show that different components of the tax system, such as the progressivity and the marginal tax rates, have had distinct and significant impacts on entrepreneurial risk-taking (see also Cullen and Gordon (2006)).[2] We contribute to the literature on taxable income elasticities (Gruber and Saez, 2002; Saez, Slemrod and Giertz, 2012), but bringing hassle costs and the quest for simplicity into the picture, and also by exploiting the heterogeneity across self-employed individuals to provide evidence for tax evasion.

Our analysis of how members of the same household jointly optimize (and misreport) their self-employed earnings echoes the analysis of the joint income decisions among wage earners in Eissa and Hoynes (2004), Eissa and Hoynes (2006), and Gelber (2014). We contribute to this literature by inferring tax evasion from households' reporting behavior far from versus close to the eligibility thresholds for the simplified and super-simplified self-employment regimes.

Our work also relates to how tax payers respond to costly information with inattention as in Hoopes, Slemrod and Reck (2017) or with behavioral biases as in Lockwood and Taubinsky (2016) and Lockwood (2016). Here again, we exploit what the data tell us regarding how individuals' learning reacts differently over time to different changes in self-employment regimes, to infer evidence of tax evasion.

[2] Gentry and Hubbard (2000) find that a progressive tax system discourages entry into entrepreneurship. Using the Panel Study of Income Dynamics, Bruce (2000) finds that reducing marginal tax rates on self-employed income reduces the probability of entry into self-employment, while reducing the average tax rate slightly increases entry.

More generally, our work speaks to the literature on the determinants of entrepreneurship (see, among others, Hamilton (2000), Schoar (2010), Adelino et al. (2015), and Schmalz et al. (2016)), but we focus specifically on the role of fiscal incentives, taxation, and administrative simplicity. Most closely related are papers on entrepreneurship in France, using other sources of administrative data. Lelarge, Sraer and Thesmar (2008) look at the effects of credit constraints on entrepreneurship using variation from a French loan guarantee program. Hombert, Schoar, Sraer and Thesmar (2017) show that unemployment insurance can stimulate self-employed activity in France. Here we go one step further by looking at the motives (tax evasion, quest for simplicity) for choosing simpler self-employment regimes.

A series of recent studies makes use of the new French administrative data. Fack and Landais (2010) and Fack and Landais (2016) study charitable contributions. French tax data also is used in two important contemporaneous papers that study income and wealth distributions in France, by Garbinti, Goupille-Lebret and Piketty (2017) and Garbinti, Goupille-Lebret and Piketty (2016). Our focus is on individuals' choice between self-employment regimes and the determinants of observed bunching patterns.

A copious literature applies the bunching methodology to a wide range of topics such as intertemporal allocation in response to mortgage contracts changes (Best, Cloyne, Ilzetzki and Kleven, 2015b), transaction taxes in housing markets (Best and Kleven, 2016), corporate taxation (Best, Brockmeyer, Kleven, Spinnewijn and Waseem, 2015a), responses to the EITC (Chetty, Friedman and Saez, 2013), the social insurance earnings test (Gelber, Jones and Sacks, 2017a; Gelber, Jones, Sacks and Song, 2017b), and fuel efficiency requirements (Slemrod and Salleel, 2012). We extend the methodology to identify tax evasion and simplicity motives for enrollment in the simplified self-employment regimes.

Finally, our work is related to the many empirical studies of misreporting in response to taxation, especially recent examples of which are Carrillo et al. (2017), Feldman and Slemrod (2007), Pomeranz (2015), and Gordon and Slemrod (2000). We contribute to this literature by exploiting the multiplicity of self-employment regimes, their staggered introduction, and the heterogeneity between individuals (including between individuals that claim the same income level) to identify the motives for misreporting.

The remaining part of the paper is organized as follows. Section 2 presents the evolving landscape of self-employment in France over the period 1994-2015, describing the various self-employment regimes and the dynamic sequence of self-employment reforms. Section 3 presents the data and provides some descriptive statistics. Section 4 provides evidence of individuals' bunching at the eligibility thresholds for the simpler self-employment regimes. Section 5 provides evidence of tax evasion as a determinant of the observed bunching. Section 6 performs a structural estimation which both, confirms the tax evasion motive and uncovers the quest for simplicity. Finally, Section 6 concludes.

3.2 The Landscape of Self-Employment in France 1994-2015

In this section, we describe the landscape of self-employment in France over the period 1994 to 2015, by providing details on the institutional background in France, the different fiscal incentives in place, and their evolution over time.

3.2.1 A Primer on the French Personal Income Tax and Social Insurance System

We start with a brief note on the French tax and social insurance system and the features that will be relevant for the self-employed.

Taxable income of a household is the sum of all the sources of income – including income from self-employed activities – minus exemptions and deductions (itemized and standard). Given that the French tax system is household-based, each household has a scaling factor called the number of parts, which is determined by the household composition. For a single adult, that scaling factor is one, for a married couple, it is 2. The first two children each add 0.5, the following children add 1 part each. Having a disabled child adds another 0.5 part. For example, a married couple with a child has a number of parts equals to 2.5. A married couple with 3 children has a number of parts equals to 4, and a married couple with one disabled child has a number of parts equals to 3.

The tax bracket cutoffs are expressed in terms of the so-called family coefficient, defined as:

$$\text{Family coefficient} := \text{FC} = \frac{\text{household taxable income}}{\text{number of parts}}$$

Appendix 3.C.2 shows how the tax liability of a household is determined. In brief, the family coefficient serves the same role as the taxable income in the U.S. for determining the tax bracket and total tax paid "per-part." To get the total tax liability of the household, the "per part" tax is inflated by the number of parts.[3]

An important feature of the tax system is that taxable income does not map directly into the tax rate: a given taxable income can lead to a wide range of tax rates based on family structure, which will be helpful in our analysis and for the estimation.

Employed and self-employed also have to make a sizable contribution (around 30%) of their earnings to the system of social security. These payments are collected and managed by entirely different government bodies than the income tax. They go towards government-provided health insurance,

[3]Figure 3.21 shows the income tax schedule for fiscal years 1994, 2006, and 2012. The tax schedule changes almost every year as part of the yearly budget voted by the French Parliament.

workers' compensation, disability insurance, social insurance and public pensions, as well family-related and means-tested transfers. For the self-employed these social insurance contributions are levied on the same base as the income tax, but with a different timing and some adjustments. Social insurance contributions depend on the type of activity. Additionally, there are many different rates and contribution schemes for agents with different professions. This contributes to the significant heterogeneity in the total tax rate (income tax plus social insurance contribution rate) faced by agents, even conditional on the same total income.

3.2.2 Self-Employed Regimes

Activity types:

For tax purposes, the self-employed are classified into three types of activities. These are important because they affect the policy parameters facing an agent, which we describe below. The three types are: (i) the "Industrial and Commercial Services" category, referred to as "I&C Services" below,[4] (ii) the "Industrial and Commercial Retail" category, referred to as "I&C Retail,"[5] and the (iii) the "Non Commercial" category.[6].

These activity types, defined for fiscal purpose, do not necessarily align well with the underlying economic characteristics of businesses. For instance, developing and selling software pertains to the Non Commercial type, while purchasing and selling equipment goods pertains to the I&C Retail category. Similarly, bakery, butchery, or restaurant businesses are counted as I&C Retail activities, while construction work, plumbing, carpentry, and auto or other repair shops and dry cleaning count as I&C Services. Moreover, all professional activities, such as consulting, private coaching, translation services, sales agents services, expert services, empty property subleasing, as well as all liberal professions (doctors, notaries, or lawyers in private practices) belong to the Non Commercial category.

Three self-employed regimes

We focus on self-employed businesses that operate under the personal income tax code.[7] As of 2009, these self-employed could choose one of three regimes:

(1) The standard regime[8]

---

[4]These are the so-called *Bénéfices Industriels et Commerciaux Services*.
[5]*Bénéfices Industriels et Commerciaux Vente*.
[6]*Bénéfices Non Commerciaux*.
[7]A self-employed individual who owns his business can also choose to incorporate and be subject to the corporate tax code. We do not study those individuals, who typically operate on a larger scale than the businesses studied here.
[8]*Régime Réel*.

(2) The simplified regime[9]

(3) The super simplified regime[10]

These three regimes can be characterized along five main dimensions summarized in Table 3.1:

1. *Eligibility requirements:* The super simplified and simplified regimes can only be chosen by agents with revenues below a threshold y^*_{kt}, which depends on the type of activity k (with k = I&C Retail, I&C Services or Non Commercial) and on the fiscal year t. Figure 3.1 shows the thresholds' evolution. The thresholds are not very high for the Services and Non Commercial activities (equal to 32,600 € in 2012), but much higher for the Retail activities (81,500 €). In the case of the super simplified regime, there is an additional condition on the family coefficient as of year $t-2$, which has to be below a year-specific threshold f^*_t that corresponds to the third tax bracket cutoff.[11] Figure 3.2 schematically represents the regime options. An agent with revenues below the threshold (y^*_{kt}) for his activity type in a given year can choose between the simplified, super simplified, and standard regimes. Above the threshold, the only possible option is the standard regime.[12] In practice and to avoid a costly regime change when an agent's revenues fall just above the threshold, a "tolerance region" is created (which is, e.g. 6.1% higher than the actual threshold in 2012 for the Services and Non Commercial Activities and 9,9% higher for the Retail Activities): the entrepreneur can remain in the simpler regimes if her self-employed income falls within this region for a maximum of two years in a row.[13] This is shown as the hatched area in Figure 3.2. If the eligibility threshold is crossed for more than two years or if the tolerance threshold is crossed, the special regime status is lost, and the agent has to file under the standard regime. In addition, certain types of professions cannot operate under the simplified or super simplified regimes, most notably agricultural activities, leasing of durables and equipment, leasing of professional or non-furnished buildings, and real estate businesses. Additional activities excluded from the super simplified regime include liberal professions such as lawyers, doctors, insurance agents, or accounting experts, and formally registered artists rewarded through copyright.

[9]We lump together under this heading two regimes: the (1) *Régime Micro-entreprise* and the (2) *Régime Auto-Entrepreneur sans Option Libératoire*. These two regimes are indistinguishable in the tax data because they differ only in the social contributions paid, which do not appear on the tax returns. These two regimes are fairly close, the social contributions being set to generate the same liability (except for sales much smaller than the thresholds), and they were merged in January 2016.

[10]*Régime Auto-Entrepreneur avec Option Libératoire*.

[11]For instance, that cutoff was 26,420 € for year 2010, so that for households to be eligible for the super simplified regime in 2012, their family coefficient in 2010 had to be lower than 26,420 €.

[12]In theory, there is a limit of 750,000 € for self-employed in the standard regime. We will not study that threshold, as it makes an agent shift between the personal and corporate income tax realms.

[13]This grace period and tolerance threshold do not apply in the first year after the business' creation.

2. *Definition of the income tax and social security base:* In the standard regime, taxable income is net business income, i.e. the difference between gross revenues and costs, including depreciation of assets and investments according to standard accounting rules. In the simplified regime, taxable income is equal to revenues times a scaling factor $1-\mu$, where the rebate μ is determined by the tax administration. It depends on the activity type and has changed over time (see Figure 3.1). In the super simplified regime, taxable income is simply revenues (i.e. the rebate $\mu = 0$).[14] In the simplified and super simplified regimes, an agent cannot claim any deficits.

3. *Income tax and social insurance contribution rates paid on the base:* In the standard and simplified regimes, the regular tax and social insurance contribution rates apply (both of which differ across households depending on several factors as explained above). In the super simplified regime, the agent pays a flat rate that simultaneously takes care of both income tax and the social insurance contributions. The flat rate differs by activity and has changed over time. It is completely unrelated to an agent's actual income tax bracket or tax rate that applies to the rest of his (non super simplified) income. Thus, even an agent in the zero income tax bracket still has to pay the flat rate times revenues for all his activities that fall under this regime. In the simplified regime, a minimal social security contribution is due even at zero revenues.

4. *Accounting requirements and tax avoidance:* Self-employed in the standard regime have to keep detailed accounts to document their revenues and costs, following standard rigorous accounting practices. Businesses in this regime can join a certified accounting center (hereafter, CAC), which helps them keep and check their accounts and serves as a guarantee of sound fiscal conduct to the tax authority. In practice, almost all join a CAC because not doing so results in the business' taxable income being inflated by 25%. Self-employed in the simplified and super simplified regimes only need to report their revenues and are not required to comply with rigorous accounting practices. They are nevertheless required to keep private accounts for their activity, as well as receipts from purchases and sales in case of an audit (much like any tax payer who would, e.g. claim itemized deductions).

5. *Tax hassle:* In addition to having to keep different types of accounts, which involve more hassle for the standard regime than the simplified ones, there are also differences in how easy it is to file taxes. In the standard and simplified regimes, tax payments occur annually at the normal tax filing date and social insurance payments occur separately through the regular social insurance procedure, thus requiring two separate filings. In the super simplified regime, tax and social insurance payments are due monthly or quarterly, based on actual realized revenues (cash in hand), and are all taken care of at the same time, thus minimizing filing and hassle.

[14] A subtlety to note is that, to determine the overall tax bracket of the household, it is the revenues times $1-\mu$ where μ is the same rebate as in the simplified regime above that is added to the rest of a household's income. It is not the full amount of revenues that is added, which would make the super simplified regime very unattractive.

In addition, the standard regime is the only one subject to the Value Added Tax: self-employed in this regime charge VAT on their products sold and claim VAT on their inputs.

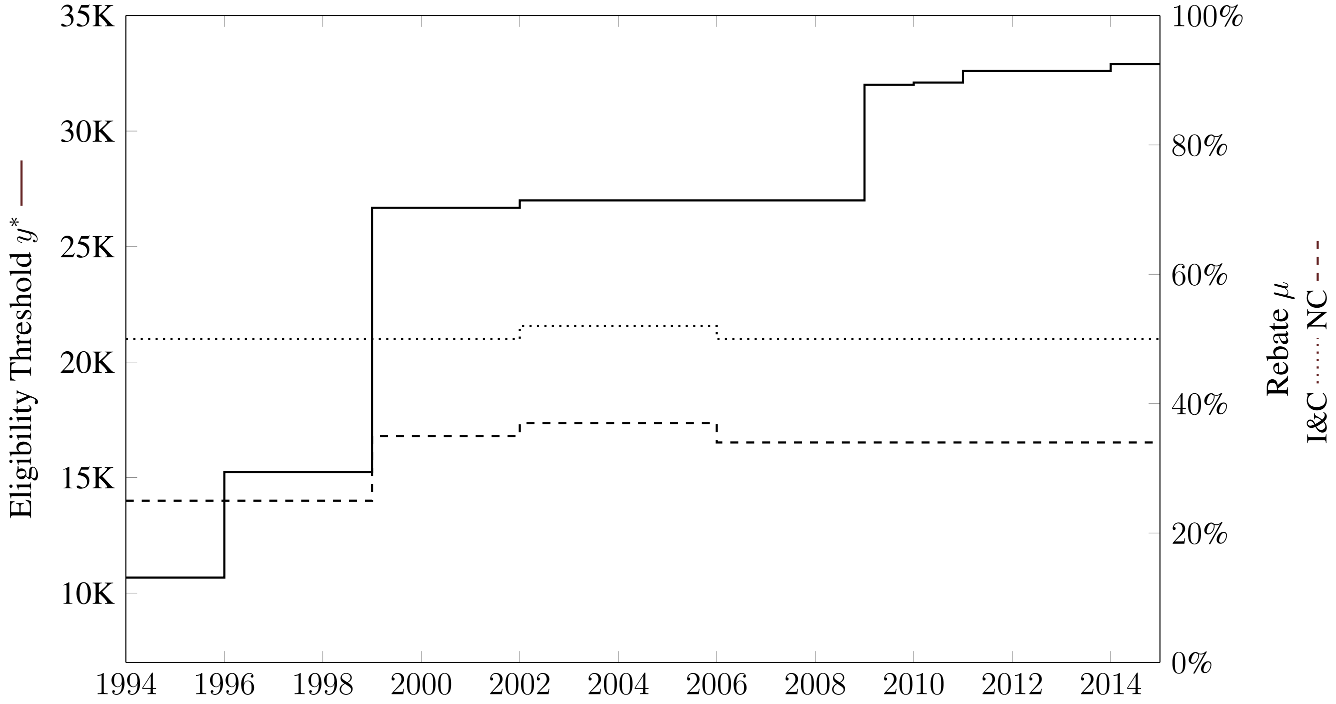


Figure 3.1: Eligibility Thresholds and Rebate for the Simpler Regimes

Notes: The figures shows the eligibility threshold in euros along with the rebates for different activity types. The threshold is identical between Industrial & Commercial services and Non Commercial activities, it is represented as a single solid line.

On balance, the key advantage of the standard regime is that it allows subtraction of input and operating costs from taxable income. This is advantageous for businesses with employees, significant investments, or high operating costs (higher than the rebate μ of the simpler regimes).[15] The main advantages of the simplified and super simplified regimes are lighter "hassle" costs (less stringent accounting and tax reporting requirement) and more scope for misreporting income (due to the substantial tax incentives it provides, the accounting is very often double-checked by a certified accounting center in the standard regime, whereas there exists no such tax incentives in the simple regimes). In addition to its being simpler than the simplified regime, the super simplified regime features a low flat rate compared with the sum of the regular income tax and social insurance contribution rates.

[15]In addition, self-employed in this regime can benefit from tax credits, such as those for R&D spending, e.g. *Crédit d'impôt recherche* or *crédit d'impôt compétitivité et emploi*, and some government help in special zones, none of which are available when filing under one of the simplified regimes.

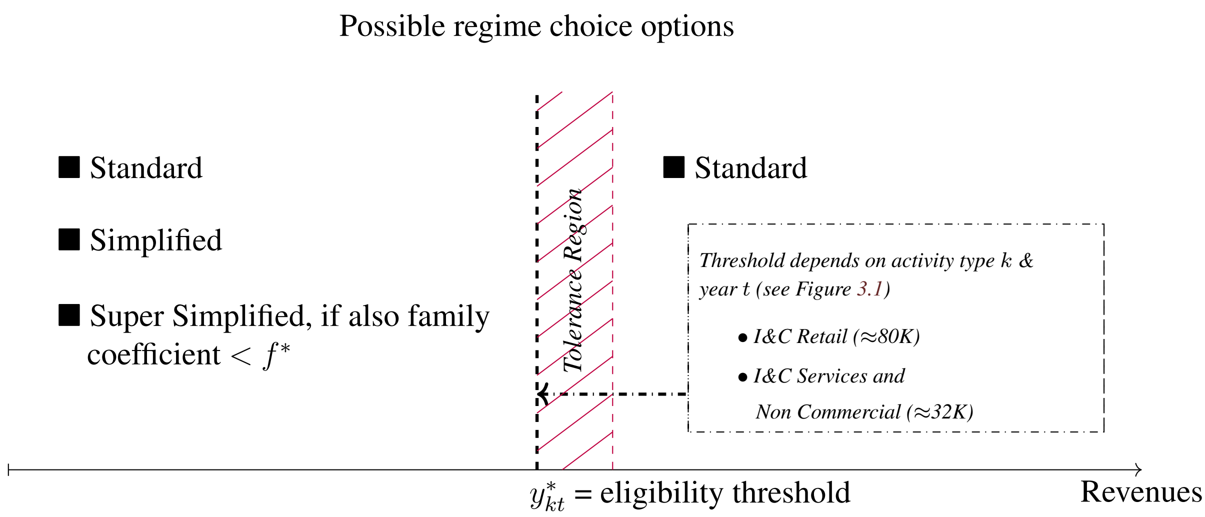


Figure 3.2: Eligibility Thresholds and Regime Choice Options

Table 3.1: Summary of the self-employed regimes

	Regime		
	(1) Standard	(2) Simplified	(3) Super simplified
Eligibility	None	Revenues $< y^*_{kt}$	Revenues $< y^*_{kt}$ + $\text{FC}_{t-2} < f^*$
Income tax & SI contribution base	Net business income	Gross revenues $\times$ (1-rebate)	Gross revenues
Income tax & SI contribution rate	Standard	Standard	Flat rate
Avoidance	Detailed	Only for audit	Only for audit
Tax hassle	Annual, separate tax and SI	Annual, separate tax and SI	Monthly or quarterly, joint tax and SI

Two key reforms:

The thresholds and rates applicable to each regime have changed extensively over time as shown in Figure 3.1. These changes generate policy variations that are key for our analysis. Two major reforms stand out. First came the 1999 reform which greatly extended the eligibility threshold for the simplified regime from 100,000 French Francs (15,244 €) to 500,000 Francs (76,220 €) for I&C Retail activities and to 175,000 Francs (26,678 €) for the I&C Services or Non Commercial activities. Before 1999, the thresholds were so low as to only apply to very small activities and the simplified regime was not a reasonable option for many self-employed. Second came the 2008 reform which created the super simplified regime. This reform stemmed from the political will to further increase tax simplicity by replacing the social insurance contributions and income taxes by a unique tax transfer proportional to self-employed revenues.

3.3 Data and Descriptive Statistics

We now describe our datasets and provide new summary statistics on the demographic and economic characteristics of the self-employed across activities, years, and regime types.

3.3.1 Data

Our main data consists of the entire tax returns of the full sample of French households (around 33 million per year) over the period 2011-2015[16] from the French Internal Revenue Service.[17] In addition, we have access to the tax returns of a representative sample of 500,000 households over the period 1994-2017. The income tax returns contain comprehensive income data at the individual and household levels, as well as key demographic information such as household composition, individual age, and gender.

We also make use of a quite unique data source, the *Enquête Revenus Fiscaux et Sociaux*, which consists of tax returns for a subsample of the population that are matched to large-scale employment survey data and benefits receipt data. This combined dataset covers the period 1996-2015 and has a sample size of around 100,000 respondents per year. It contains some highly useful variables such as education, type of profession and occupation, social insurance benefits and government transfers received, standard of living, and tax free capital income.

[16]There are around 1200 variables.
[17]*Direction Générale des Finances Publiques (DGFiP).*

Finally we use the business register when analyzing the effects of reforms on the creation of new self-employed businesses and on the switches of individuals between the various regimes.[18]

3.3.2 Descriptive Statistics on Income and Demographics of the Self-Employed

These new datasets generate some key original summary statistics on the demographic characteristics and incomes of the self-employed in France.

Key characteristics of the self-employed

Figure 3.3 shows the evolution over time of the fraction of self-employed among the labor force population aged 15 to 64. The two vertical red lines represent the 1999 and the 2008 reforms. We distinguish between two groups of self-employed individuals: those earning self-employed income, possibly earning incomes in other categories as well (in red) and those earning only self-employed income (in blue). The fraction of self-employed has remained stable over time until 2009, at around 6.7%, whereas the fraction of self-employed with only self-employed income has been slightly decreasing from 5.4% in 1994 to 4.8% in 2009. Self-employment as a whole has risen since 2009, and the upward trend is even larger for the number of individuals earning self-employed income on top of a salary income. In 2015 the fraction of individuals with at least some self-employed income had risen to 8.4%.

Table 3.2 shows the demographic (in Panel A), income (in Panel B) and income tax (in Panel C) characteristics for the following three subgroups: (i) pure wage earners, (ii) pure self-employed, (iii) agents who earn both self-employed and wage income. Averages are taken over the full sample period 1994-2015. The self-employed tend to be on average 8-9 years older than pure wage earners and almost three times as likely to be retired. Contrary to what could be expected, the self-employed are not particularly young. There is no significant difference between those three groups regarding the presence or number of children, yet the self-employed are more likely to be married or in a civil union. Despite being, by definition, employed or earning self-employed income for at least part of the year, a significant portion of individuals have also claimed unemployment benefits at some point during the year; however, the self-employed are less likely to have done so than pure wage earners.[19] The variable “Educated ”is equal to one for agents who have at least a professional or training high-school level degree, but not the highest high school level degree (the *Baccalauréat*).[20] There is no

[18]The datasets are *Fichier de comptabilité unifié dans SUSE* (FICUS) and *Fichier approché des résultats d’Esane* (FARE).

[19]The interplay between unemployment insurance and self-employment could inform the studies on the design of unemployment benefits and their insurance value as in Landais, Kolsrud, Nilsson and Spinnewijn (2015) and Landais and Spinnewijn (2017).

[20]In French, the two professional high school level degrees are the *Certificat d’Aptitude Profesionnelle (CAP)* or the *Brevet d’Études Professionnelles (BEP)*.

Figure 3.3: Evolution of Self-Employment 1994-2015

Notes: "Self-Employed" are individuals who earn any self-employment income (and may or may not also receive additional wage income). "Self-Employed only" are individuals who only earn self-employed income. The numbers are the fraction of self-employed or self-employed only among the labor force population aged between 15 and 64. The two red vertical lines represent the 1999 and the 2008 reforms.

significant difference in the fraction of self-employed and wage earners with a high-school degree, but the self-employed are significantly more likely to have completed at least a bachelor-level university degree. The variable “High skill ”identifies agents that are in higher skilled occupations, such as licensed professionals, teachers, engineers or executives of the public service or private sector. The self-employed are significantly more likely to be employed in high skill occupations.

Panel B shows that the individuals who receive self-employed income only earn on average around 33,000 € per year. Those who earn self-employed income in addition to salaried income have on average 30,000 € of self-employed income and only around 6,000 € of wage income per year. Self-employed agents have more than three times as much capital income as wage earners (around 6,000 as compared to 1,900). The same goes for tax free capital income. The variable “standard of living ”measures the total disposable income per adult equivalent at the household level. On average, those with at least some self-employed income have a 25% higher standard of living than pure wage earners.

To form groups consistent over the years of individuals with comparable tax rates, we define four tax brackets. For each sub-period 2011-2013 and 2014-2015, we first group individuals having a 0 tax rate into the tax bracket 0, and then cluster the other individuals according to their average tax rate into three groups. Table 3.3 describes these four groups. For instance, over 2011-2013, the tax bracket 1 comprises individuals with an average tax rate strictly positive and below 5.9%, the tax bracket 2 those with a tax rate between 5.9% and 13.5%, while the tax bracket 3 contains all individuals with average tax rates higher than 13.5%. Panel C of Table 3.2 shows the distribution across tax brackets for each group. Self-employed individuals are more than three times as likely to fall in the highest tax brackets relative to pure wage earners.

Tables 3.8, 3.9 and 3.10 reproduce Table 3.2 for different periods. Doing so reveals several interesting findings. First, wage income adjusted for inflation has been consistently rising, but average self-employed income experienced a fall post 2008, at the same time as the number of agents with self-employed income rose (Figure 3.3). We return to this in more details below when we study the impacts of the 2008 reform introducing the super simplified regime. Second, capital income increased significantly for all groups. Third, the proportion of self-employed who perceive unemployment benefits at some point during the year doubled from the 1999-2008 to the 2009-2012 period. Such an increase did not occur for wage earners.

Differences between Industrial and Commercial and Non Commercial activities

Table 3.4 shows the demographic and socioeconomic variables for the self-employed split by type of activity. We first see a significant gender gap: women are significantly more represented in Non Commercial activities, although they are underrepresented among the self-employed in general. Retirees concentrate in the I&C Retail and I&C Services activities. The most educated and highly skilled self-employed are found in the Non Commercial activities, which includes professions such as physi-

Table 3.2: Summary Statistics: Self-Employed and Wage Earners 1994-2015

	All	**With wage income only**	**With self-employed income only**	**With any self-employed income**
Panel A				
Age	40	40	49	48
Female	0.47	0.48	0.32	0.33
Married or in Civil Union	0.50	0.49	0.63	0.62
Has any children	0.41	0.41	0.39	0.41
Number of Children	0.71	0.71	0.70	0.72
Retired	0.06	0.06	0.17	0.14
Claimed unemployment benefits	0.11	0.11	0.03	0.05
Claimed any social insurance benefits	0.48	0.48	0.38	0.39
Educated	0.72	0.72	0.73	0.76
Bachelor (at least)	0.15	0.15	0.21	0.24
High skill	0.12	0.11	0.19	0.20
Panel B				
Wage Income	19576	20549	0	6005
Self-employed Income	2004	0	32982	29934
Capital Income	2154	1875	5148	6047
Tax free capital income	1161	1072	2467	2351
Standard of living	42607	41845	50208	53312
Panel C				
Zero tax bracket	0.16	0.16	0.15	0.14
Low tax bracket	0.32	0.33	0.23	0.22
Medium tax bracket	0.38	0.39	0.31	0.32
High tax bracket	0.14	0.13	0.31	0.32
Panel D				
Population (in mill.)	532.7	497	26.3	35.6

Notes: "With wage income only" refers to individuals with only wage income. "With self-employed income only" refers to individuals with self-employed income only. "With any self-employed income" refers to people with any self-employment (who may also have wage income). "All" refers to any individual who earns either some wage income or some self-employed income (or both). All variables in panels A and C are expressed in percent of the full group represented in each column. "Age" is expressed in years. "Number of Children" is the average number of children. Panel B provides average income in each category in constant 2012 euros. Panel D gives the total population in each column for the whole period. See Section 3.3.2 for the definition of all variables.

Table 3.3: Average Tax Rates of the individuals constituting the Tax Brackets

	2011-2013				2014-2015			
Tax Bracket	Obs	Min	Mean	Max	Obs	Min	Mean	Max
0 (Zero)	1,223,256	0	0	0	1,094,301	0	0	0
1 (Low)	667,888	0.01	3.30	5.91	488,810	0.01	3.92	6.71
2 (Medium)	659,258	5.92	8.53	13.46	439,805	6.72	9.51	14.34
3 (High)	245,668	13.47	18.41	≈ 45	181,765	14.35	19.19	≈ 45

Notes: This table provides the description of the average tax rates of the individuals constituting the various tax brackets. For example over the sub-period 2011-2013, 667,888 individuals belong to the Low tax bracket, with average tax rates ranging from 0.01% to 5.91%.

cians in private practice, lawyers or notaries. Perhaps as a result, self-employed with Non Commercial activities have a higher standard of living and higher capital and labor incomes.

Differences between regimes

Table 3.5 shows the same summary statistics separating the regimes. The first two columns compare the standard and simplified regimes, which existed throughout the full sample period; the last three columns compare the simplified, standard and super simplified regimes over the period 2009-2012. Agents in the simplified regime are on average older and far more likely to be retired. Agents in the super simplified are on average younger, less likely to be married, and much more likely to be claiming unemployment and social insurance benefits. Agents in the simplified and super simplified are less educated and in lower skill occupations than those in the standard regime.

Entrepreneurs in the standard regime claim nearly four times as much self-employment income as those in the simplified or super simplified regimes. Despite having somewhat lower wage revenues, agents in the standard regime therefore enjoy a significantly higher standard of living (i.e. disposable income per adult equivalent) than other self-employed, and belong thus more often to higher tax brackets.[21]

[21] Many entrepreneurs in the simpler regimes use self-employment as a complement to wages or pensions. Self-employed with large businesses requiring their full work capacity would typically choose the standard regime.

Table 3.4: Self-Employed Earners by Type of Activity 1994-2015

	All	**Industrial and Commercial (Retail and Service)**	**Non Commercial**
Panel A			
Age	48	49	46
Female	0.33	0.28	0.41
Married or in Civil Union	0.63	0.65	0.59
Has any children	0.41	0.39	0.44
Number of Children	0.73	0.68	0.80
Retired	0.14	0.16	0.11
Claimed unemployement benefits	0.05	0.05	0.05
Claimed any social insurance benefits	0.40	0.39	0.41
Educated	0.76	0.67	0.90
Bachelor (at least)	0.24	0.10	0.49
High skill	0.20	0.08	0.43
Panel B			
Wage Income	6049	5265	7538
Self-employed Income	30505	22718	45376
Capital Income	6133	6040	6552
Tax free capital income	2303	1997	2790
Standard of living	53642	45317	69444
Panel C			
Zero tax bracket	0.13	0.16	0.08
Low tax bracket	0.22	0.26	0.14
Medium tax bracket	0.32	0.34	0.29
High tax bracket	0.33	0.24	0.49
Panel D			
Population (in mill.)	34,7	22.5	12.6

Notes: “Industrial and Commercial (Retail and Services)” refers to the group of agents who have some revenues from I&C Retail or I&C Services activities. “Non Commercial” refers to the group of agents who have some revenues in a Non Commercial Activity. “All” refers to agents who have some self-employed income in at least one of the two aforementioned categories. See the notes to Table 3.2.

Table 3.5: Self-Employed Earners by Regime

	1994-2008		2009-2012		
	Standard	**Simplified**	**Standard**	**Simplified**	**Super-Simplified**
Panel A					
Age	46	52	48	50	43
Female	0.30	0.34	0.34	0.38	0.37
Married or in Civil Union	0.68	0.58	0.61	0.52	0.44
Has any children	0.47	0.29	0.43	0.30	0.36
Number of Children	0.84	0.50	0.76	0.52	0.62
Retired	0.07	0.31	0.10	0.28	0.13
Claimed unemployment benefits	0.02	0.07	0.03	0.10	0.21
Claimed any social insurance benefits	0.40	0.33	0.41	0.38	0.54
Educated	0.77	0.68	0.83	0.76	0.81
High skill	0.22	0.15	0.26	0.19	0.15
Panel B					
Wage Income	3945	10439	4470	10868	7985
Self-employed Income	39446	11522	40925	11848	10307
Capital Income	5938	6174	7864	6713	2484
Tax free capital income	2294	2452	2464	2611	1032
Standard of living	56814	42434	66278	47553	39086
Panel C					
Zero tax bracket	0.09	0.19	0.08	0.18	0.23
Low tax bracket	0.20	0.28	0.16	0.24	0.27
Medium tax bracket	0.29	0.32	0.37	0.40	0.42
High tax bracket	0.42	0.22	0.39	0.18	0.08
Panel D					
Population (in mill.)	19.3	7.3	4.6	3.1	0.9

Notes: "Standard" refers to agents in the standard regime; "Simplified" to agents in the simplified regime; "Super simplified" to agents in the super simplified regime. See the notes to Table 3.2.

3.4 Bunching in the Simpler Regimes

In this section, we provide graphical evidence of individuals' behavioral response at the eligibility thresholds. The eligibility thresholds are notches, at which several things change at the same time: i) the tax base, and therefore the tax liability (inclusive of social insurance liabilities), ii) the hassle cost, and iii) the ease of evasion. However two agents reporting the same self-employed revenue can face widely disparate tax incentives depending on their activity type (which affects operating costs and the rebate) and their total income and family situation (as explained in Appendix 3.C.2, agents with the same self-employed and total income can lie in different income tax brackets because of the peculiarity of the French tax system).

Only agents who would choose to be in the simpler regime and earn revenues above the eligibility threshold absent the notch face a discontinuity in their tax schedule. These are for instance agents with low operating costs and high hassle costs. For agents who do not want to be in the simple regimes there is no discontinuity at the thresholds. Thus, when evaluating the excess mass of revenues, we should not take into account agents with revenues below the eligibility threshold who choose the standard regime.[22]

In our analysis, we focus mostly on the I&C Services and Non Commercial activities, because the threshold for the I&C Retail activities is so high that the income distribution density at that level becomes very thin. We do use the I&C Retail activity group as a placebo group around the (much lower) threshold applicable for the I&C Services and Non Commercial activities (where the distribution of I&C Retail income is sufficiently dense).

In the graphs where we pool all self-employed agents (in all regimes), we need to work with taxable income z (rather than revenues y). This is because the self-employed in the standard regime have to report only their net income (revenues minus costs); those in the simplified and super simplified regimes only have to report their revenues (and not their costs). We convert these revenues into the common and comparable (reported) "taxable "income z as defined above. Similarly, we convert the thresholds for the simplified regime into a taxable income equivalent.[23] In the graphs with only agents in the simplified or super simplified regimes, we can directly work with revenues y.

[22]Since we do not actually see revenues for agents in the standard regime, when we plot graphs in terms of revenues, we are automatically taking into account only agents who are in the simple regimes and ignoring those who choose the standard regime.

[23]This yields two thresholds (one for I&C Services equal to $z^*(1 - \mu_{\text{I\&C Services}})$ and one for Non Commercial activities ($z^*(1 - \mu_{\text{I\&C Retail}})$), due to their different rebates. The thresholds for the super simplified regime is the same whether expressed in revenue or taxable income since $z_f = y_f$ and $\mu_{\text{Super simplified}} = 0$ in that regime, regardless of the activity.

3.4.1 Regime Choice Responses: Graphical Evidence

The choice between one of the simpler regimes (simplified or super simplified) and the standard regime depends on the combination of all factors described in Section 3.2, and is thus complex. An agent would tend to choose one of the simpler regimes if her operating costs and investment are low, if she has no employee and finds tax hassle particularly burdensome, or if she wants to evade taxation.

We start by checking whether the proportion of agents who choose one of the simpler regimes (simplified or super simplified) exhibits uneven behavior around the eligibility thresholds. Figure 3.4 shows in 2014-2015 the fraction of agents in any simplified regime among all self-employed, as a function of their taxable income. The vertical lines represent the eligibility thresholds expressed in taxable income for, respectively, the simplified I&C Services, the simplified Non Commercial, and the super simplified (I&C Services and Non Commercial together). The fraction of agents in a simple regime spikes at all those thresholds – which we see as a first evidence for a behavioral response to the thresholds.[24]

3.4.2 Income Responses: Graphical Evidence

We now turn our attention to the income responses to the notches created by the eligibility thresholds. We focus on the 2011-2015 period, for which we have exhaustive data on income tax returns and over which policy parameters have remained fairly stable.

Method

We can identify the excess mass in self-employed revenues as illustrated in Figure 3.5 to the left of the eligibility threshold. Recall that except for revenues that fall in the tolerance region during the grace period (as described in Section 3.2 and Figure 3.2), we do not observe revenues in the simpler regimes for agents above the eligibility threshold. Contrary to what is common in the bunching literature, we cannot use the empirical distribution at the right of the notch to estimate a counterfactual distribution absent the notch or a missing mass. We adapt Chetty et al. (2011)'s counterfactual and fit a flexible polynomial to the empirical distribution of income, excluding a range R_1 of income at the left of

[24] The response appears stronger at the super simplified threshold than at both simplified thresholds. This is essentially a base effect: at half the super simplified threshold, the only agents impacted have a Non Commercial activity in the simplified regime and the agents not impacted are in the simplified I&C Services and all agents in the super simplified or standard regimes. At taxable income around the super simplified threshold, there are no more agents in the simplified regime, the agents impacted belong to the super simplified regime while the not impacted belong to the standard one. The pool of individuals for which the threshold is irrelevant is bigger for the simplified I&C Services than for the simplified Non Commercial activities, itself bigger than for the super simplified. Moreover the super simplified regime features stronger bunching, as we show below.

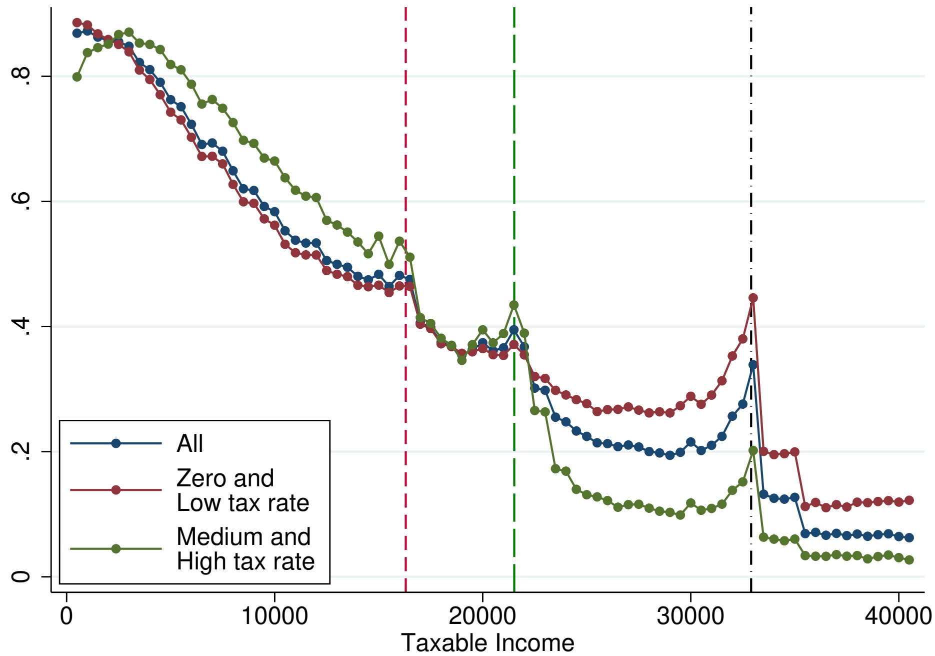


Figure 3.4: Regime Choice – Share Choosing the Simplified or Super Simplified Regime

Notes: The figure represents the share of self-employed agents who choose either the simplified or the super simplified regime among those eligible to at least one of these two regimes, based on the full population for 2014-2015. Each dot represents the average proportion of agents in a bin. The horizontal axis represents self-employed taxable income z. The red dashed vertical line represents the eligibility threshold for the I&C Services in the simplified regime (converted into a taxable income equivalent, i.e. multiplied by one minus the rebate μ applicable to I&C Services), the green dashed vertical line represents the threshold for the Non Commercial activities in the simplified regime (also converted into a taxable income equivalent), and the blue dash-dot vertical line represents the threshold for the super simplified regime for both I&C Services and Non Commercial Activities.

the threshold T. Revenues are centered around the eligibility threshold (which varies over time), and grouped by bins of 500 €: the bin B_j contains all individuals with self-employed income in the interval $]B_j - 500, B_j]$, so that all individuals reporting sales at exactly the threshold belong to B_T.

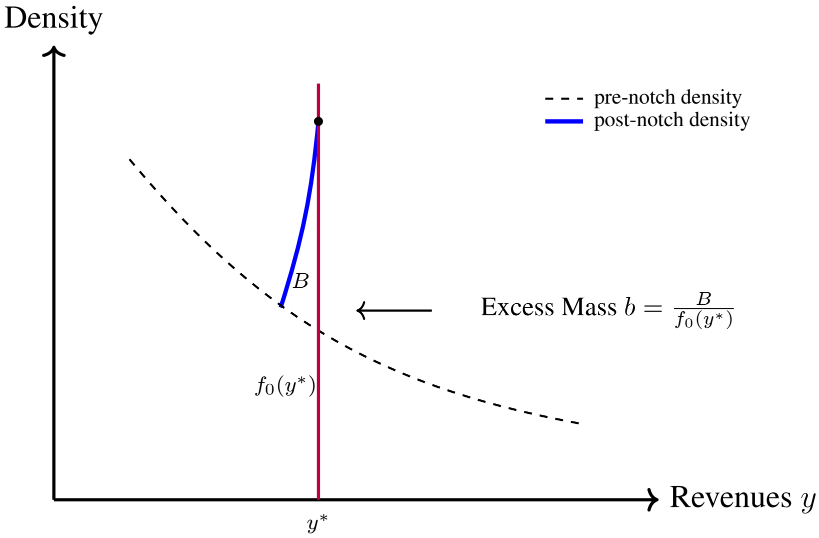


Figure 3.5: Excess Mass Method

The figure illustrates how we measure the excess mass. The red vertical line represents the eligibility threshold y^*. The excess mass B is computed by taking the difference between the post-notch density (the light blue solid curve) relative to the pre-notch density represented by the dashed blue curve (and captured by a counterfactual smooth fitted polynomial). The excess mass measure is $b = B/f_0(y^*)$ where $f_0(y^*)$ is the counterfactual density right before the threshold.

To estimate a counterfactual distribution to the left of the threshold, we fit a smooth polynomial by running the following regression:

$$C_j = \sum_{i \in A} \beta_i (B_j)^i + \sum_{r \in Round} \delta_r \mathbf{1}_j^r + \varepsilon_j \tag{3.1}$$

with j such that $B_j \leq T - R_1$, and where C_j stands for the count of individuals in income bin B_j and $A \subset \{1.5, 2, 3, 4, 5, 6, 7\}$ is the set of polynomial exponents. As in Kleven and Waseem (2013), we control for round-number bunching by adding dummies $\mathbf{1}_j^r$ equal to 1 if bin B_j contains a multiple of a round number r. Beyond improving the counterfactual distribution, this control avoids counting as bunching what is just round-number bunching when the bunching area contains round numbers. We have experimented with different round numbers, and given the size of the bin the counterfactual distribution turns out relatively similar in all cases – to the extent we include a dummy for multiples of 6,000 €, which corresponds to a monthly multiple of 500 €.[25] $Round$ is thus the smallest set of round numbers we need and it comprises the number 6,000.

We then use estimates from (3.1) to obtain the predicted counterfactual (including in the excluded range):

$$\hat{C}_i = \sum_{j \in A} \hat{\beta}_j (B_i)^j + \sum_{r \in Round} \hat{\delta}_r \mathbf{1}_j^r$$

The excess number of individuals B, capturing the number of agents in excess of what would be expected without the notch, is calculated as:

$$B = \sum_{i \in S} (C_i - \hat{C}_i)$$

where $S = \{i \in \mathbb{N} \mid B_i \in]T - R_2, T]\}$ and $R_2 \leq R_1$. We then define the excess mass b as the excess number of individuals in the bunching zone divided by the counterfactual number of individuals in the bin with upper bound $B_T = 0$, which is $\hat{C}_T$.

$$b = \frac{B}{\hat{C}_T} \tag{3.2}$$

A visual inspection of the distribution suggests the bunching behavior begins to be noticeable at 1,500 or 2,000 € from the threshold, depending on the regime, activity, sub-period or group considered. So we choose $R_1 = 1500$ or $R_1 = 2000$ depending on the situation, and $R_2 = 1000$, which provide a conservative estimate.[26]

To compute standard errors for the excess mass b (as well as for the elasticities which we compute based on B in Section 3.6), we generate earnings distributions and excess mass estimates by resampling the residuals in (3.1). The standard error is taken to be the standard deviation of the distribution of estimates generated by bootstrapping.

Results

Figure 3.6 shows in blue the distribution of self-employed revenues for agents in the simplified and super simplified regimes, pooled over all years 2011-2015. The revenues are centered around the

[25]The figures shown in the next section visually confirm that with bins of 500 € the irregularities in the empirical distribution mainly come from this round number.

[26]The results are qualitatively robust to different choices for the bunching zones. Yet while choosing a wider zone has only a limited impact on the counterfactual or the excess mass, choosing a smaller bunching zone should bias the excess mass towards 0, because of a higher counterfactual (R_2 too small) and/or the exclusion of some bunchers of the bunching area (R_1 too small).

threshold for I&C Services and Non Commercial activities (the red vertical line) and grouped in 500 € bins. The blue curve represents the count in each bin of individuals in the I&C Services and Non Commercial Activities, the red curve does the same for bins of individuals in the I&C Retail activities. The counterfactual distribution, in red, fits well the empirical distribution (and captures the round number bunching visible at 30,000 and 24,000 €). The excess mass is colored in yellow.

The sharp spike in the income distribution right before the threshold translates into an excess number of individuals estimated at 152% of the counterfactual number of individuals in the threshold bin. This bunching behavior is massive and it is precisely estimated. The placebo group, constituted of individuals in the same simple regimes but in the I&C Retail activities, and for which the eligibility threshold is much higher, shows no bunching at all.

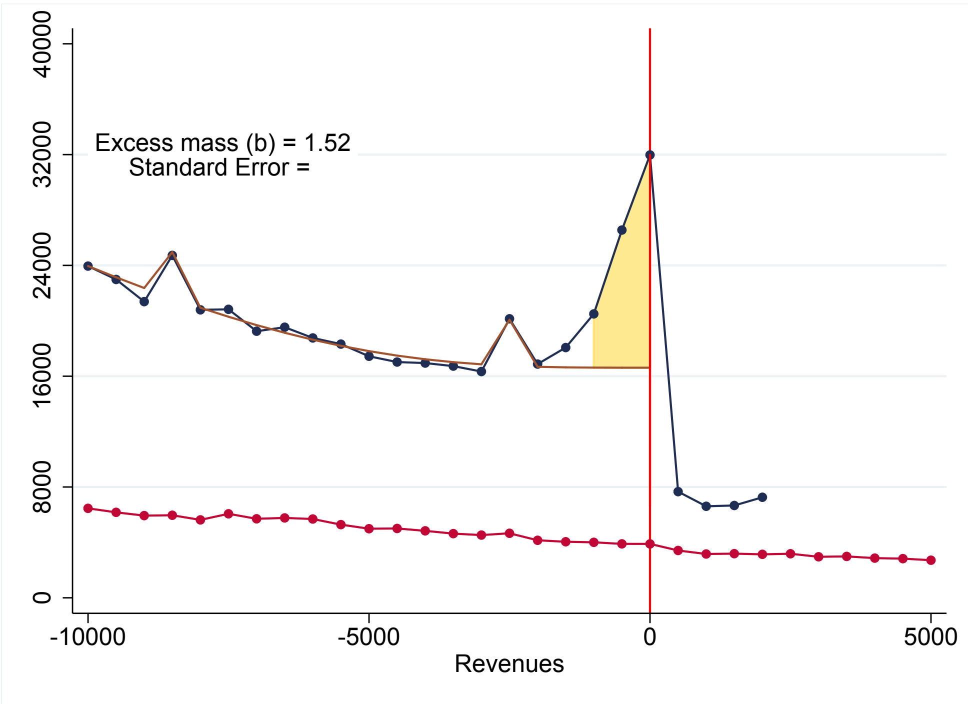
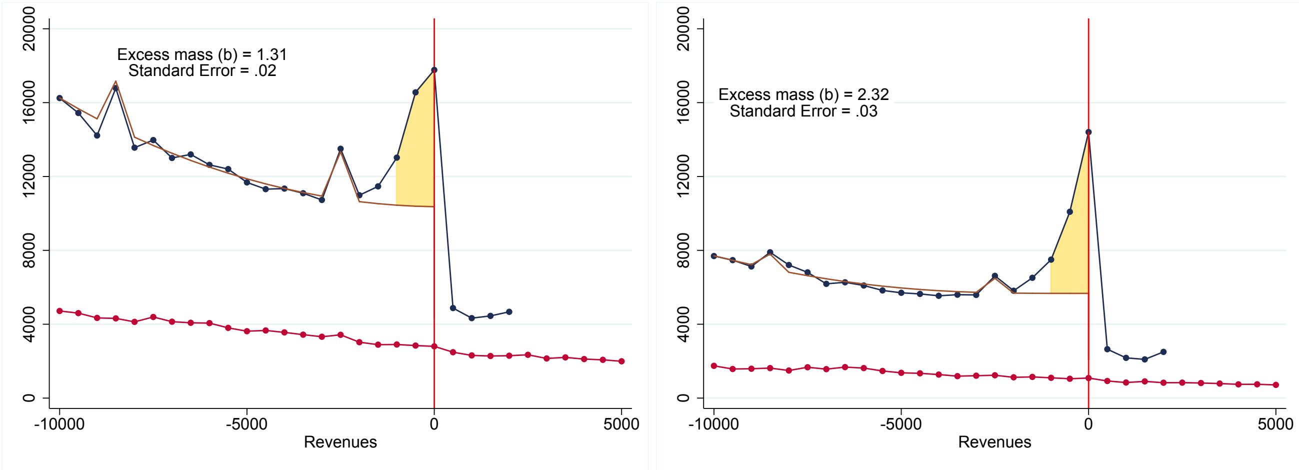


Figure 3.6: Bunching at the Eligibility Thresholds

Notes: The figure represents the frequency of revenues, by bins of revenues centered around the eligibility threshold (the red vertical line). We pool data for 2011-2015 and for all agents in the simplified and super simplified regimes. The counterfactual distribution is represented by the red solid curve in the figures and is fitted using a smooth polynomial as explained in Section 3.4. The estimated excess mass is in yellow. There is significant bunching, equal to 152% of the average counterfactual frequency within 2,000 € of the notch. The thick red curve serves as a placebo test: it shows the frequency distribution for the I&C Retail activities, centered around the eligibility threshold for the I&C Services and Non Commercial activities (the actual threshold for the I&C Retail activities is higher).

Figure 3.7 splits the sample further into agents in the simplified (panel (a)) and agents in the super simplified regime (panel (b)). The bunching at the eligibility threshold is starker in the super simplified regime, equal to 2.3 times the counterfactual density right before the threshold. Again the agents in I&C Retail activities, which serve as a placebo group, exhibit no excess mass.



(a) Simplified

(b) Super Simplified

Figure 3.7: Bunching in the Simplified and Super Simplified Regimes

Notes: Panel (a) considers bunching among agents in the simplified regime only. Panel (b) considers agents in the super simplified regime only. See the notes to Figure 3.6.

Heterogeneity in Bunching

Self-employed respond to tax incentives. Any cost or benefit advantage from being in one regime over the other is amplified for agents in higher tax brackets. Table 3.6 shows the bunching mass (b) and its standard error for agents in different tax brackets, activities, fiscal regime, and over two sub-periods 2011-13 and 2014-15 over which the policy parameters remain constant. To compute bunching by tax bracket, we exploit the peculiar feature of the French tax system that leads to vastly different tax brackets even given the same self-employed income (i.e. around the eligibility threshold). All bunching masses are statistically significant and large. Agents clearly respond to tax incentives as the excess mass increases with the tax brackets, whatever the period, activity or fiscal regime considered. We can also notice that bunching is more pronounced in the Non Commercial activities than in the I&C Services.

Agents with additional income sources. We can also measure bunching for individuals with additional sources of income. Figure 3.8 distinguishes between self-employed agents who also claim some salary income (panel (a)) and those who do not (panel (b)), as well as self-employed individuals who claim pension income (panel (c)) and those who do not (panel (d)). Self-employed individuals with also

Table 3.6: Excess mass by tax bracket

Panel A: Simplified

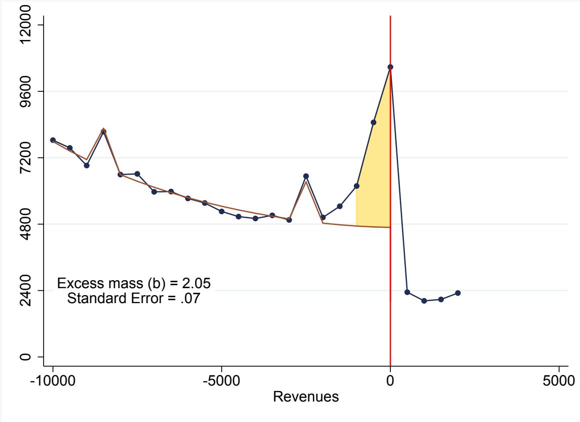
	I&C services		Non commercial	
Tax Bracket	2011-2013	2014-2015	2011-2013	2014-2015
0 (Zero)	0.63 (0.08)	0.65 (0.08)	0.77 (0.12)	0.95 (0.09)
1 (Low)	0.82 (0.08)	0.83 (0.09)	1.05 (0.07)	1.26 (0.10)
2 (Medium)	1.11 (0.07)	1.28 (0.10)	1.23 (0.09)	1.88 (0.11)
3 (High)	1.45 (0.20)	1.46 (0.24)	3.15 (0.17)	3.6 (0.27)

Panel B: Super Simplified

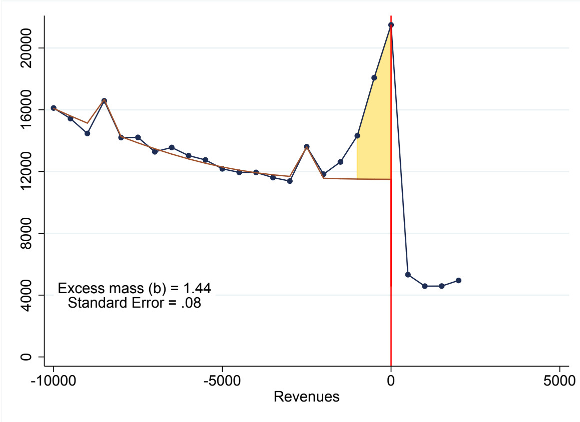
	I&C services		Non commercial	
Tax Bracket	2011-2013	2014-2015	2011-2013	2014-2015
0 (Zero)	1.41 (0.07)	1.65 (0.10)	2.65 (0.19)	2.64 (0.13)
1 (Low)	1.86 (0.19)	2.44 (0.27)	2.7 (0.36)	3.06 (0.51)
2 (Medium)	2.54 (0.31)	2.69 (0.28)	3.02 (0.26)	3.24 (0.26)
3 (High)	5.39 (0.77)	3.88 (0.64)	4.39 (0.31)	3.82 (0.44)

Notes: The table shows the excess masses b defined in equation (3.2), dividing entrepreneurs according to their fiscal regime, their activity type, their tax bracket and two sub-periods over which the threshold remains constant. Standard errors in parentheses.

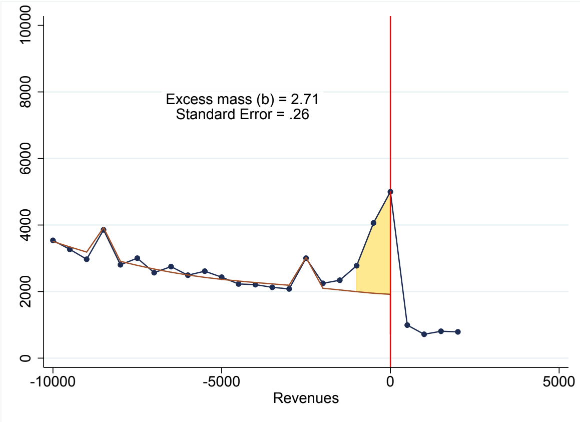
some salary income appear far more sensitive to the threshold and their excess mass is almost twice as large as that of self-employed who have no other source of income. Similarly, individuals who are retired are more than three times as sensitive to the threshold as those who are not retired. Agents with other sources of income may have higher income on average, hence presumably bigger tax incentives to respond to the threshold. They may also have a better understanding of the system, and for retirees more time to assimilate the tax system's mechanics. Finally, the availability of additional income may give more flexibility for the individual to optimize on her self-employed income, since she does not have to fully depend on it for her living. Put differently, agents with additional income have better outside options and thus are more elastic to financial incentives and/or more sensitive to hassle costs when engaging in self-employed activities.



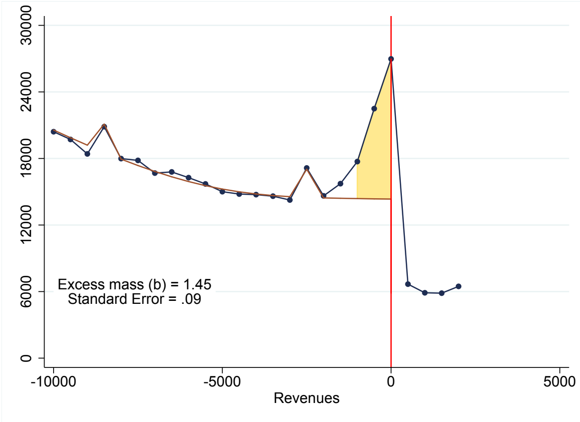
(a) With additional wage income



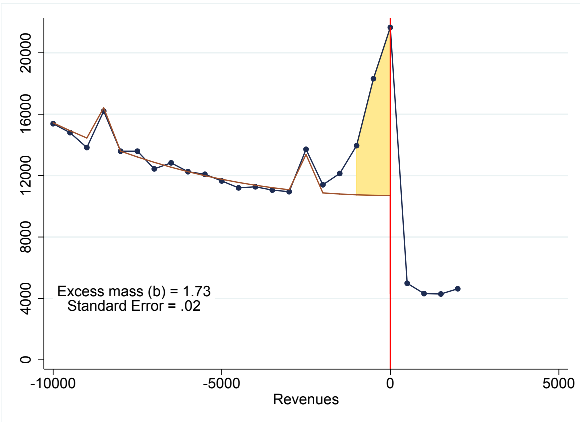
(b) Without wage income



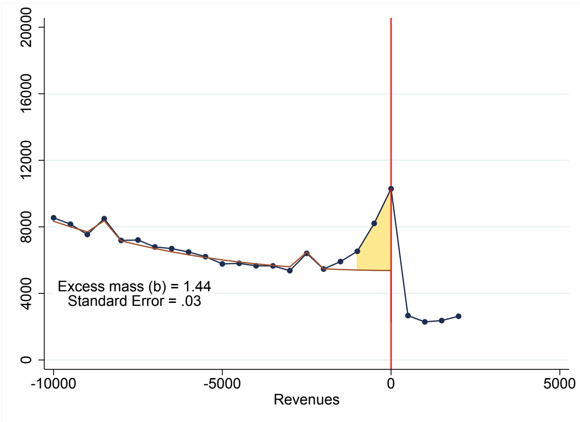
(c) With retirement (pension) income



(d) Without pension income



(e) Male



(f) Female

Figure 3.8: Bunching for Agents with or without Additional Income Sources

Notes: Bunching among different groups of self-employed agents: individuals who earn some wage income in addition to self-employed income (panel (a)) and those who do not (panel (b)), individuals with pension income (panel (c)) and those without (panel (d)), by sexe (panel (e) and panel(f)). See the notes to Figure 3.6.

3.5 Evidence of Misreporting

Do the observed bunching responses reflect a strategic choice of economic activity or tax evasion? While it is likely that part of the excess mass comes from individuals stopping their activity until the end of the year when they reach the threshold or planning their revenue streams so that they remain below the threshold at the end of year, many clues point to tax evasion as well. Although it is very challenging to prove that a given entrepreneur misreports her income, we highlight three aggregate behaviours as evidence that agents may be misreporting their income at least to some extent. The first one exploits the observation that an individual forging her revenues is attracted to round numbers. The second behaviour refers to income shifting within households. The third behaviour relates to the increase in bunching over time at the thresholds, suggesting that individuals learn how to better evade taxes over time. Before uncovering these patterns, we explain the reasons that may push a self-employed to misreport her income.

3.5.1 Why Is It Tempting to Evade Taxes in the Simpler Regimes ?

Beyond the financial incentives, the simpler regimes require fewer administrative tasks and proofs of sound fiscal accounting. In the simpler regimes, the agent does not have to send the checks of the purchases or sales she makes, she just reports total sales. In the standard regime, the tax requirement is the same but a large share of agents sends this detailed accounting to a certified accounting center, which verifies the accounting and can contact the tax authority directly if it wants to. The financial incentives to join a CAC (the deductibility of memberships and accounting expenses and avoiding a 25% inflation of the tax base) have led a large share of agents in the standard regime to join such centers. Figure 3.9 shows that at business income levels relevant for our analysis (between 15,000 and 35,000 €), almost all agents in the standard regime are CAC members. A government report (Cour des Comptes, 2014) states that conditional on an audit, the size of penalties among non-CAC members is larger than among CAC members of comparable size (around 26,000 € versus 7,000 €). It adds that the discrepancy between taxes due and taxes actually paid comes more often from genuine accounting mistakes and delays in payments and less often from outright tax evasion among CAC members than among non-CAC members.

In the end, cheating appears much easier to do in the simplified regimes, and it seems likely that among the large share of agents in the standard regime who are members of a CAC, only a minority of them intentionally misreported their revenues.

Furthermore, timing frictions in the simple regimes may induce underreporting. Agents have to decide by February of fiscal year t which regime they want to be affiliated with for their income earned in fiscal year t; however it is only at the end of year t and according to the accumulated revenues by then that individuals will know whether or not they actually qualify for the regime to which they applied. Agents with sales above the threshold can be tempted to underreport revenues to appear right

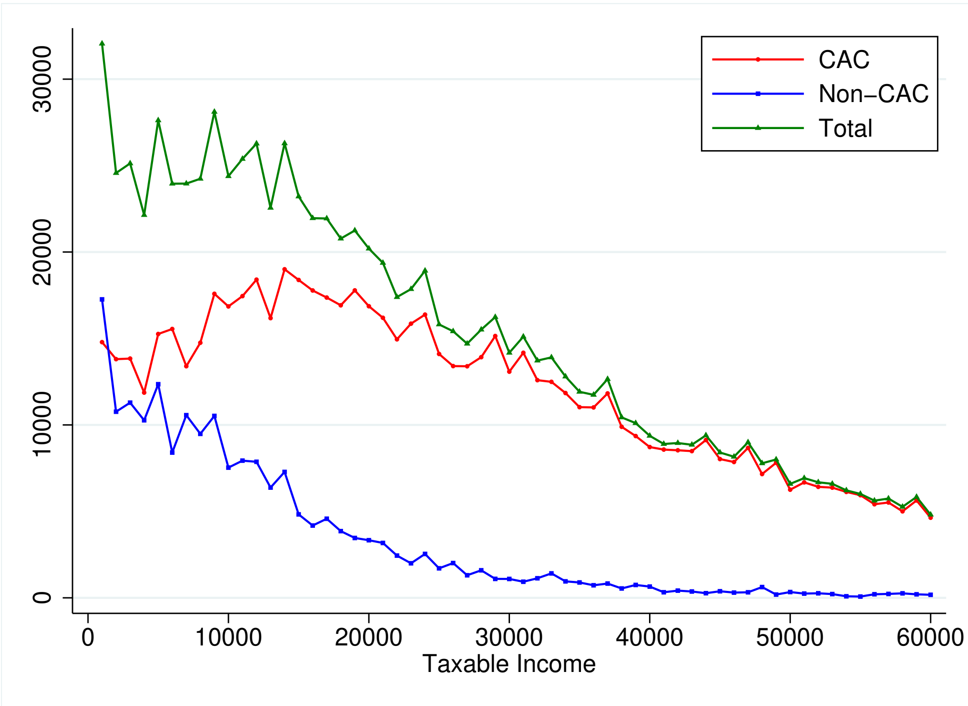


Figure 3.9: Agents in the Standard Regime Affiliated with a Certified Accounting Center (CAC)

Notes: The figure shows the number of agents in the standard regime who are members of a CAC and the number of those who are not, as well as the total number of agents in the standard regime. The figure is based on the 2011-2015 full population data. Each dot represents the average number of agents, by year. The x axis represents taxable income in the standard regime, i.e. net business income. At low income levels, there is a sizable fraction of agents who are not CAC members. That fraction declines rapidly and converges to zero at around 30,000 €.

below the threshold, as it would be very costly to have to switch status at this point.[27] In particular, a retroactive switch into the standard regime for an individual who was not part of a CAC initially entails a 25% increase in her tax base (and the non-deductibility of the memberships and accounting expenses), because an entrepreneur joining a CAC after the fifth month of the year is not entitled to these advantages. In addition in the standard regime the individual becomes subject to VAT. And if she has not kept the checks of her expenses, she may not be able to deduce the VAT spent. Finally the individual may want to save the time and energy needed to get familiar with the standard regime.

3.5.2 Round Number Bunching

The tax code makes it easier for the self-employed to misreport in the simpler regimes than in the standard regime. And in practice individuals in both the simplified and super-simplified regimes appear to use this reporting flexibility. Using the representative sample of 500,000 households, Figure 3.10 shows the share of revenues below 30,000 € that are a round number: either a multiple of 10 Francs before 2000, or a multiple of either 10 euros or 500 Francs converted into euros after the adoption of the euro in 2001. Different items from the tax returns are shown: income from each simplified regime or the standard regime, wages, capital income, as well as a random distribution (either uniform or from a gamma fit of the revenue distribution of the simplified regime). In the simplified regimes, round numbers turn out far more frequent than we would expect if the figures were randomly drawn: since the mid 2000s, 45% of reported sales are a round number where a random distribution would signal a little more than 10%. By contrast for revenues from the standard regime or for other items such as wages or capital income, the fraction of round numbers is in line with that of a random draw.

Figure 3.11 breaks down these excess round-number income by round number, focusing on revenues from the simplified regime. Multiples of 6,000, 1,000 (that are not a multiple of 6,000), 100 (that are not a multiple of 1,000) or 10 (that are not a multiple of 100) are considered (in Franc before 2000 or in euro after 2001). We also allow for round numbers in Francs after the introduction of the euro: if a reported number in euro that is not a multiple of 10 € corresponds to a multiple of 500 (1,000 or 6,000) Francs, converted to euro with the conversion rate of 6.56 or 6.55957 Francs per euro, and then rounded at the integer just below or above, this reported number in euro is counted as a round number in Francs. For each round number, we subtract from the observed number of income levels which are a multiple of this round number the number that would arise if the distribution were uniformly random. Figure 3.11 thus plots only the share of excess round numbered income. For instance in 2010 1.00% of the revenues from the simplified regime are a multiple of 6,000 € *in excess* of what we would expect if these revenues were uniformly distributed (about 1/6,000). First, "rounder"

[27] The tolerance region can limit the misreporting, but only for revenues below the tolerance threshold and for a maximum of two years. The self-employed may also fear the signal that this choice may send to the tax authority, or not be aware of the existence of this facility.

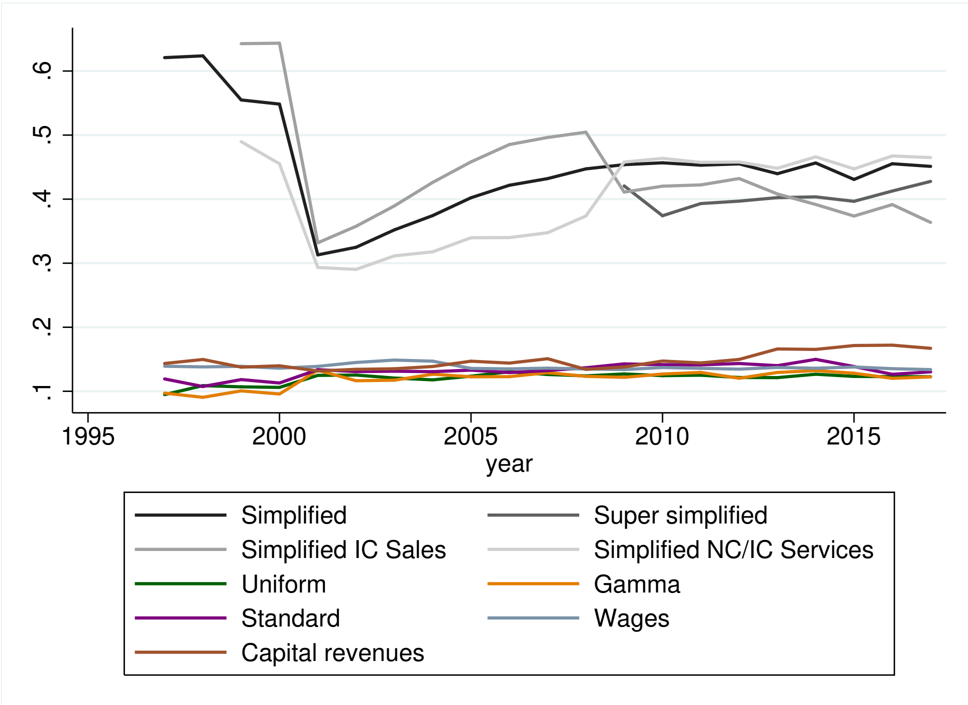


Figure 3.10: Share of round-numbered declarations, by type of revenue

Notes: This Figure shows the share of revenues below 30,000 € that are a round number: either a multiple of 10 Francs before 2000, or a multiple of either 10 euros or 500 Francs converted into euros after the adoption of the euro in 2001. Different items from the tax returns are shown: income from each simplified regime or the standard regime, as well as the breakdown between I&C Services and Non Commercial activities on the one hand and I&C Retail activities on the other hand for the simplified regime, wages, capital income, as well as a random distribution (either uniform or from a gamma fit of the revenue distribution of the simplified regime). Data from the representative sample of 500,000 households.

numbers (6,000 or 1,000) are, relative to their natural occurrence with a random draw, more often reported than "less round" numbers (10 €). Second, more rounding occurs in the Franc era. Yet in 2001 when the euro is introduced, the count of *euro*-round numbered incomes is precisely coherent with a random draw, but the count of *Franc*-round numbered incomes substantially exceeds what is expected (by almost 20% of the revenue statements). In practice, this means that we observe income streams for instance at 18,293 € much more often than expected by pure chance – they correspond to 120,000 Francs –, but as many income streams at 18,000 € as expected. The excess of *Franc*-round incomes are visible until at least 2005-2006 and slowly fade away thereafter. At the same time, *euro*-round incomes in excess steadily increase and account for a third of the revenue statements at the end of period, versus virtually none in 2001. Ultimately the share of round revenues that would not be round under a random distribution increases from less than 20% in 2001 to a third in 2008 and stabilizes at that level thereafter. The same breakdowns at comparable income levels show no meaningful excess round-numbered tax filings, whatever the round number, in the standard regime and for wages (Figure 3.22).

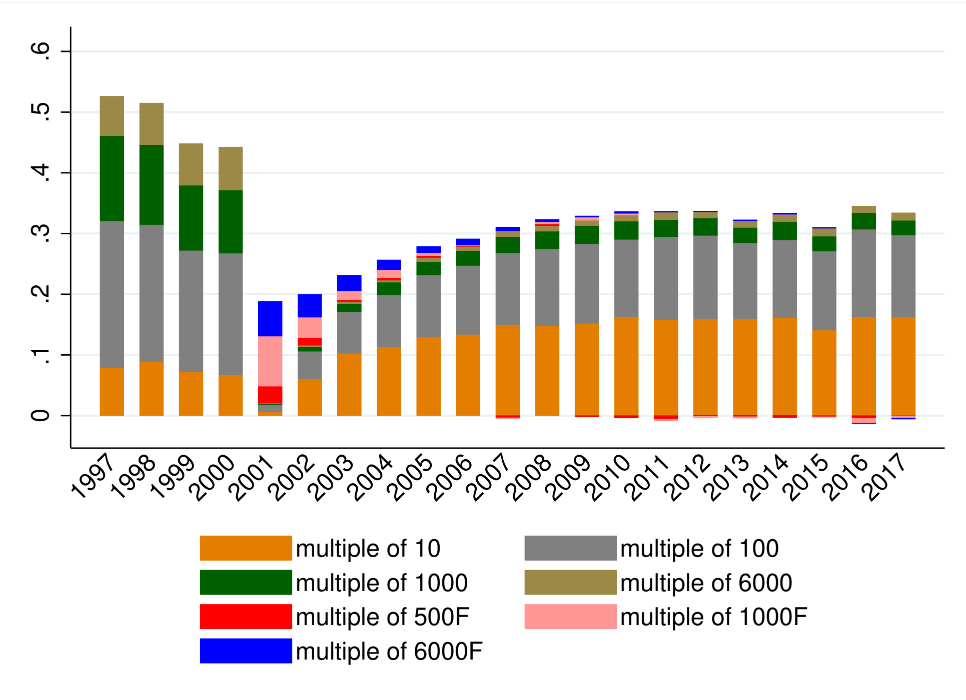


Figure 3.11: Excess share of declarations in the simplified regime, by round number

Notes: This Figure plots, for each round number, the observed share of income levels which are a multiple of this round number minus the share that would arise if the distribution were uniformly random. After 2001, a multiple of 6,000 Francs is a number that is not a multiple of 10 € and corresponds to the integer above or below of a multiple of 6,000 Francs converted into euros with the conversion rate of 6.55957 or 6.56. This Figure uses only revenues from the simplified regime below 30,000 €. Data from the representative sample of 500,000 households.

However the observation of excessive round numbers in the simpler regimes is not enough to prove evasion. Individuals might have simply decided to round up (or even down) their revenue statements just for convenience, the discrepancies between true and reported amounts may then turn out small, and this practice – though illegal – is not what comes to mind when thinking about evasion. Berson et al. (2019) highlight the role of inattention and show that individuals often round their wages when answering a survey. In a similar setting as ours but for Pakistan, Kleven and Waseem (2013) find that taxpayers tend to report round-numbered taxable income. They observe this phenomenom mainly for self-employed and only marginally for wage earners. Given that wage earners' income is mostly third-party reported, they interpret it as poor record keeping among self-employed. Another possibility is that entrepreneurs in the simplified regimes disproportionately have only round contracts. Yet we conjecture that close to the threshold, some of the agents in excess underreport their sales to be allowed to file under a simpler regime. We have in mind the political economy literature showing fraud in elections: this is indeed another instance where agents report forged numbers. In particular Kobak et al. (2016) show that in all Russian elections since 2004, "the number of polling stations reporting turnout and/or leader's result expressed by an integer percentage (as opposed to a fractional value) was much higher than expected by pure chance". To highlight the difference between the amount of rounding far and close to the threshold, we fit the distribution of reported incomes with a polynomial and dummies accounting for round-number bunching, and we allow for a different value of some of these dummies in the bunching area and outside of it. We split the exhaustive dataset between two sub-periods during which the threshold remained unchanged (2011-13 and 2014-15) and we estimate equation (3.3) separately for the two sub-periods.

$$\begin{aligned}\ln(n(x)) = \quad & P(x) + \alpha_1 \mathbf{1}_{6000} + \alpha_2 \mathbf{1}_{1000} + \alpha_3 \mathbf{1}_{10} \\ & + \beta_1 \mathbf{1}_{100} \mathbf{1}_{\text{pre bunch area}} + \gamma_1 \mathbf{1}_{100} \mathbf{1}_{\text{bunch area}} \\ & + \beta_2 \mathbf{1}_{25} \mathbf{1}_{\text{pre bunch area}} + \gamma_2 \mathbf{1}_{25} \mathbf{1}_{\text{bunch area}} + \varepsilon_x\end{aligned} \tag{3.3}$$

where $n(x)$ is the number of revenues reported at each euro x in a given activity and a given fiscal regime and $P(x)$ is a polynomial in x with exponents from $\{-4 \dots 5\}$. The regression includes dummies indicating whether x is a multiple of 6,000, 1,000 (but not 6,000), 100 (but not 1,000), 25 (but not 100) or 10 (but not 50), or whether x belongs to the bunching area defined as the 2,000 € just below the threshold. The difference between the coefficients γ_1 and β_1 (γ_2 and β_2) capture the increase in round-number bunching for multiples of 100 € (respectively 25 €) between the bunching area and below this area. The regression uses reported income in the various regimes and categories between 17,000 € and the threshold.

This specification first provides a magnitude of the round-number bunching observed at all income levels for the simpler regimes. Figure 3.13 reports the estimates of coefficients α_1 (green triangle), α_2 (red cross) and α_3 (blue circle) for both the Industrial and Commercial Services and Non Commercial activities on the one hand and the retail activities on the other hand, in the super simplified and the simplified regimes, as well as all activities together in the standard regime. Round number bunching is much more common in the simplified and super simplified regimes than in the standard one. For example in the I&C Services and Non Commercial activities of the simplified regime, a multiple of 6000 appears $\exp(4.6) = 100$ times more than a similar number ending with 1-9. All these coefficients are estimated clearly much higher in both simpler regimes than in the standard regime, with the log-scale magnifying the differences.[28] These coefficients remain fairly constant over the two sub-periods.

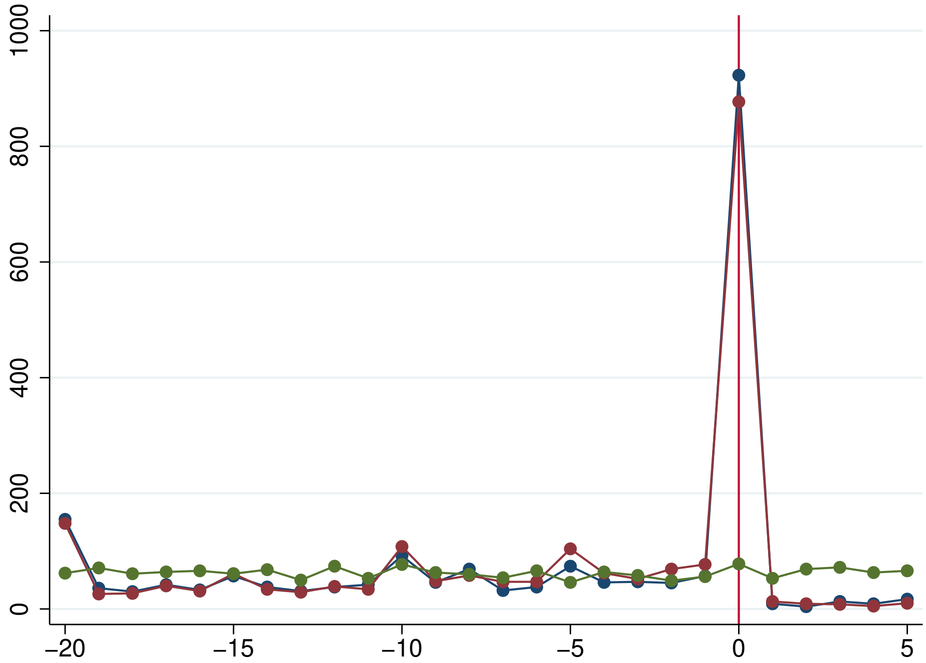


Figure 3.12: Bunching at the Exact Eligibility Threshold

Notes: This figure plots the number of individuals with income at each euro between -20 and +5 € around the eligibility threshold, over 2011-2015 and in the simplified regime (in blue), the super simplified regime (red) or the standard regime (green).

Specification 3.3 tests whether the round number bunching behavior differs close to the threshold. For the same sub-periods, activities or regimes as detailed in Figure 3.13, Figures 3.14 and 3.15 show the estimates of (β_1, γ_1) and (β_2, γ_2). For both sub-periods and both simple regimes in the I&C

[28]To help visualize the spikes in the distribution associated with a round number, Figure 3.12 plots the count of revenues for each euro between -20 and +5 € around the eligibility threshold. The sharp spike visible at the exact euro of the threshold mixes a mild round number bunching effect (since the threshold is at 32,600 or 32,900 € during 2011-15) and a salience/reference effect. Nothing is visible for agents in the standard regime, because i) the threshold has no meaning in this regime (no salience/reference impact), and ii) round number bunching is much less visible in this regime.

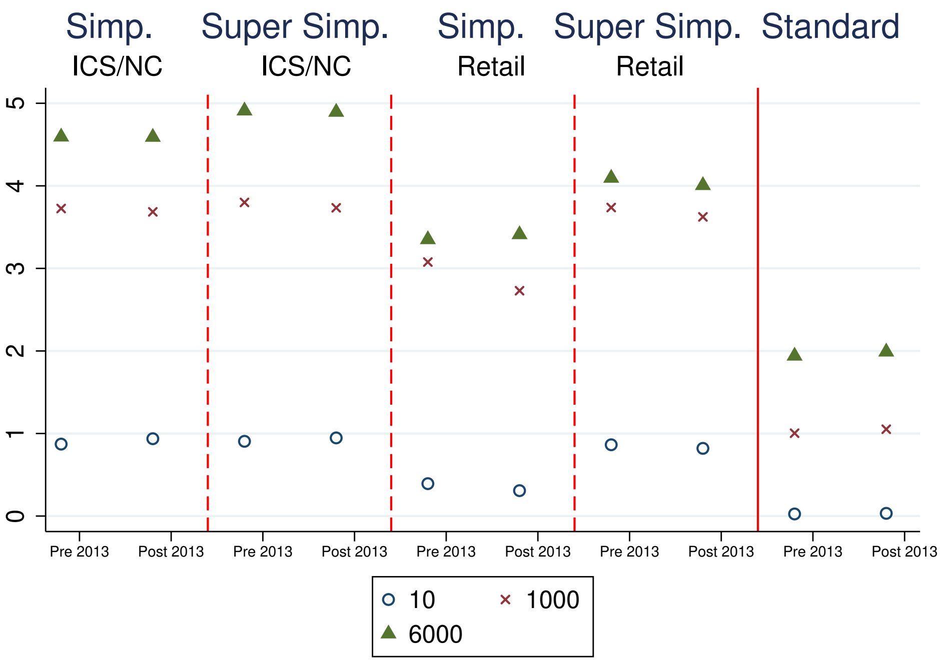


Figure 3.13: Excess number of declarations at round numbers

Notes: This graph plots the coefficient estimates α_1, α_2 and α_3 from equation (3.3). Different regressions are estimated for the two sub-periods (2011-13 and 2014-15), and for the simplified regime (Industrial & Commercial Services and Non Commercial activities on the one hand versus Retail activities on the other hand), the super simplified regime (same breakdown by activity) and the standard regime (all activities together).

Services and Non Commercial activities, agents report more often multiples of 100 or 25 € in the bunching area than outside the area. However, this is not the case for the retail activities in these simple regimes or for the standard regime, for which the threshold does not apply. In other words, the data do not follow the same data generating process in the bunching area than outside of it, but only for the regime and activities subject to the threshold. This increased rounding in the bunching area is a first indication that some of the agents in excess in the bunching area forge their revenue statements. We describe in the next section additional behavioral patterns specific to the bunching area pointing in the same direction.

Figure 3.14: Excess number of multiples of 100 (inside vs outside bunching zone)

Notes: This graph plots the coefficient estimates β_1 (red circle) and γ_1 (orange cross) from equation (3.3). See the notes of Figure 3.13.

3.5.3 Income Shifting Within the Household

The eligibility thresholds apply at the individual level; if there are two self-employed agents in the same household, each has to remain below the eligibility threshold in order to be in the simplified or super simplified regime. This provides a potential incentive to shift income within the household, i.e. to attribute some of the self-employed income earned by one person to the other one. Of course, even at the individual level, the mere presence of the eligibility threshold provides some incentive to

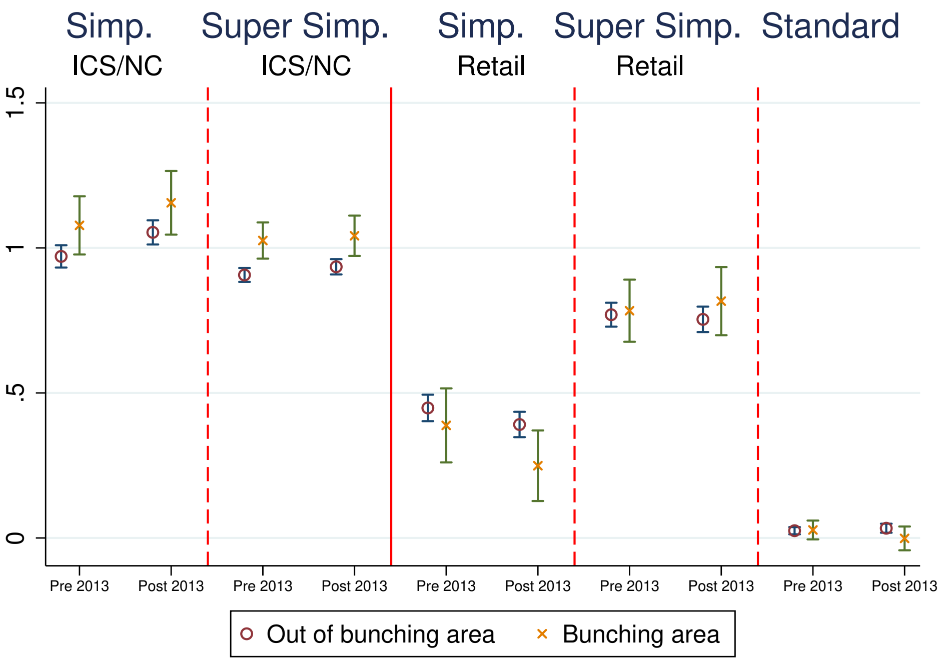


Figure 3.15: Excess number of multiples of 25 (inside vs outside bunching zone)

Notes: This graph plots the coefficient estimates β_2 (red circle) and γ_2 (orange cross) from equation (3.3). See the notes of Figure 3.13.

misreport income. However, it may be easier for a household with two self-employed earners to shift income from the taxable base of one to that of the other than it is for a single self-employed person to misreport. Agents may have a sense that reporting the income in *some* tax form field, even if not the correct one, is less reprehensible than not reporting it at all (i.e. outright evading).[29]

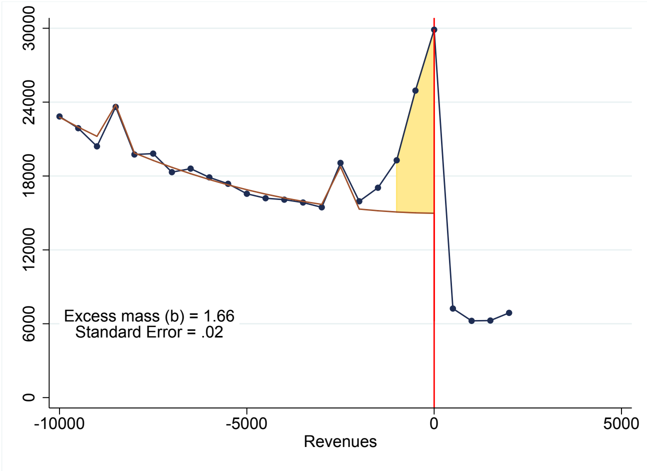
Figure 3.16 shows the excess mass for two groups of agents: individuals in households of at least two people, but who are the only one self-employed earner in a simplified regime of the household (panel (a)) and individuals in households with two self-employed members in a simplified regime (panel (b)). We will call the two household members "partners" for simplicity, although they are not always spouses or civil partners. Agents with a partner also self-employed in a simple regime exhibit significantly larger excess masses, as one may expect if it is easier to report income below the threshold by being also able to shift income between the two partners (in addition to all other possible response margins such as real responses and misreporting). Other interpretations are possible; for instance people in households with two self-employed earners may be better informed about the thresholds and more willing to misreport (individually). But we do not think this is the main explanation.

Panels (c), (d) and (e) offer more direct evidence for shifting. They focus on the sub-population of households with two self-employed members in a simplified regime. Panel (c) plots the revenues of the partner with the smallest self-employed revenues among the two partners against the revenues of the partner with the largest revenues. The revenues of the lower earner increase markedly as the higher earner's revenues start approaching the threshold. A similar increase in the earnings of the lower earner occurs when the higher earner approaches the upper bound of the tolerance region. One possible interpretation of this pattern is that, as the higher earner becomes threatened to no longer be eligible for the simplified regimes, he shifts some of his revenues to his partner.[30]

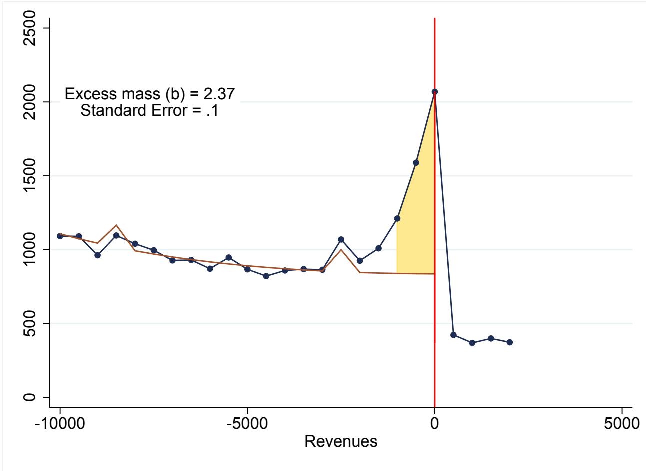
Panel (d) plots the distribution of agents in a simplified regime with a partner with self-employed income within 100 €. The excess mass is very large: while it is uncommon for two partners to report about the same self-employment income in general, it becomes more common very close to the threshold. Finally panel (e) plots the distribution of the sum of the revenues of the two self-employed household members relative to twice the eligibility threshold, represented by the red vertical line. We observe an excess mass at twice the eligibility threshold. We compare this to the bunching by "placebo partners". To build placebo partners, we separately take the earners from the households with two self-employed earners and randomly match them to another self-employed earner in a simplified regime.

[29] Furthermore, if household members actually help each other in their respective self-employed activities, they may even perceive this income shifting as somewhat fair, if not entirely legal.

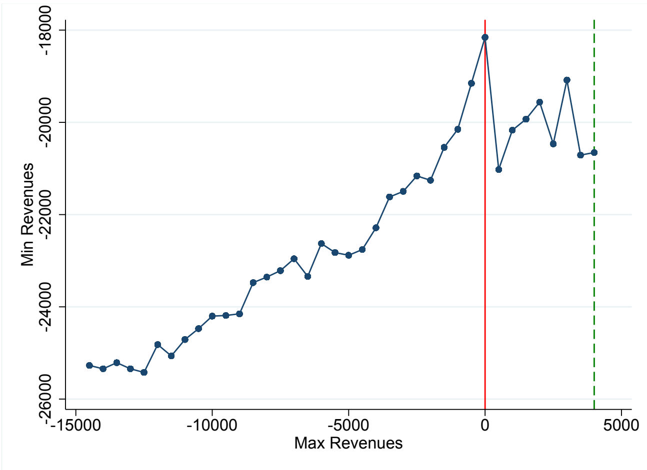
[30] The argument here relies on the increase in the secondary earner's revenues being sharp enough around the eligibility or tolerance thresholds. Otherwise, it may just be the result of selection or assortative matching.



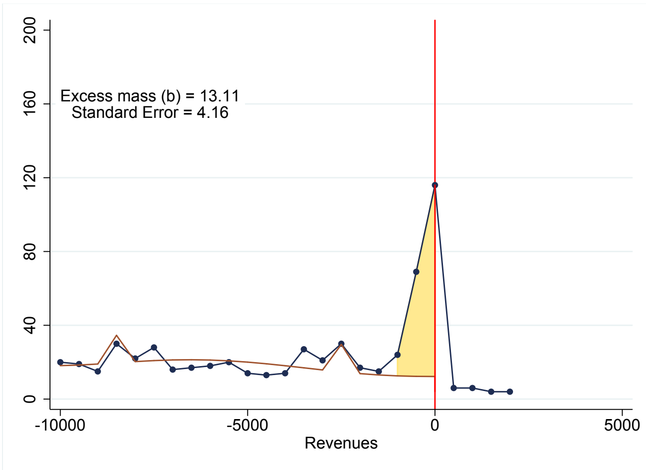
(a) In households with one self-employed agent



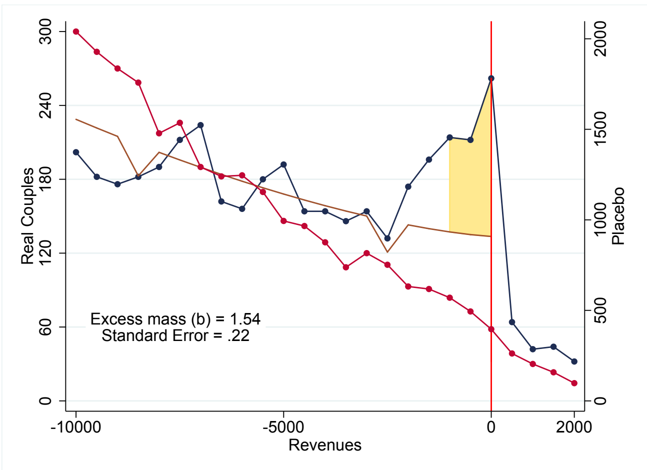
(b) In households with two self-employed agents



(c) Lower earner's revenues



(d) Same revenues



(e) Sum of revenues

Figure 3.16: Income Shifting Within the Household

Notes: Panel (a) (respectively, Panel (b)) shows bunching among individuals in households with two (respectively, one) self-employed earner(s) in a simplified regime. Panel (c) shows the centered revenues (relative to the eligibility threshold) of the lower earner in a household with two self-employed earners in a simplified regime against the centered revenues of the higher earner. Panel (d) shows the bunching for the higher earner in households with two self-employed earners, conditional on the lower earner's simplified revenues standing at less than 100 euros from the higher earner's. Panel (e) plots the sum of revenues of two partners in the same household who are both in a simplified regime; in this panel only, the red vertical line is at twice the individual eligibility threshold. The red line shows the distribution of the sum of revenues

We then plot the sum of the revenues within the placebo partners. There is no bunching at twice the threshold.[31]

In the end, some households very likely optimize their joint revenue statement by shifting some income between themselves so that both can report income below the threshold (and by maybe not reporting at all the excess joint revenue).

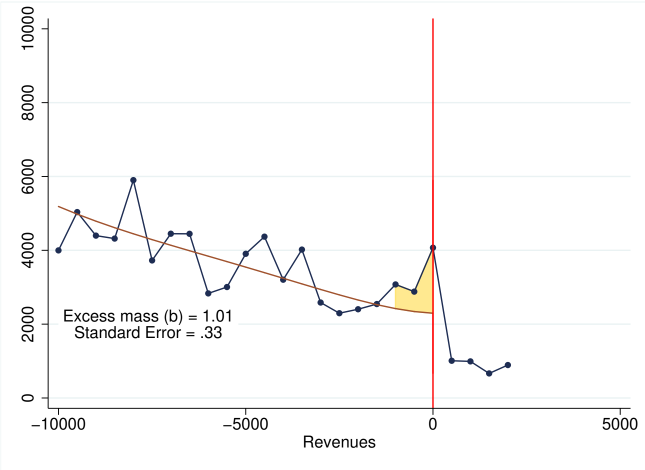
3.5.4 Learning to Evade

Saez (2010) shows that the bunching at the first kink of the Earned Income Tax Credit increases over time. Chetty et al. (2013) add that individuals moving to higher-bunching area respond more to the EITC (they learn), whereas individuals moving from high-bunching area to low-bunching area continue to respond to the EITC (they already know how the system works and do not forget). Consistent with the view that individuals learn slowly over time, Figure 3.17 shows the dynamics of bunching from the extension of the simplified regime in 1999 to the introduction of the super-simplified regime in 2009. Panel (a) shows the bunching at the eligibility threshold on average over the period 1999-2001, panel (b) over 2002-2005, and panel (c) over 2006-2008. Bunching clearly increases over time. Figure 3.18 shows the same increase over time for income shifting within couples.

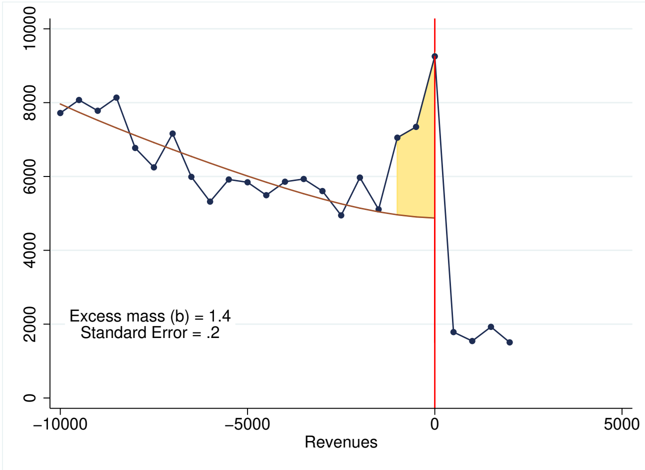
The costly learning hypothesis predicts that agents react rapidly and massively to a large and salient change, whereas they should respond only slowly to a small or complex change. Larger changes require lower attention costs to be noticed, and the financial losses from not adapting to them are larger. We point to three events where individuals' response (or lack of response) is fully consistent with this hypothesis.

First the 1999 reform is an example of a massive and easy to understand change in the tax system: it did not introduce any new feature, but the eligibility threshold was multiplied by five for the I&C Retail activities and by almost two for the I&C Services and Non Commercial activities. Figure 3.19 plots the number of self-employed agents (panels (a) and (c)) and their average self-employed income (panels (b) and (d)), broken down by fiscal regime and normalized to their 1994 value (panels (a) and (b)) or 2004 value (panels (c) and (d)). Panels (a) and (b) show that the responses to the 1999 reform were immediate and took place mostly during the first year: the incentives appear to have been sufficiently clear and large. Since this reform did not change the overall tax simplicity of being self-employed, there was only a very modest increase in the overall number of self-employed: the sharp increase of 60% in the number of agents in the simplified regime in 1999 was counterbalanced by

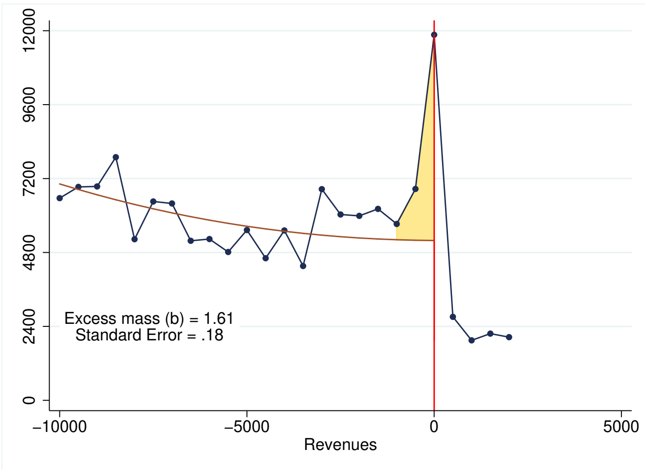
[31]To check that this pattern is not driven by the fact that households with two earners in a simplified regime are more prone to bunching at the individual threshold in general, we also check that if we form placebo partners by randomly matching only people from within two self-employed earner households, there is no bunching either.



(a) 1999-2001



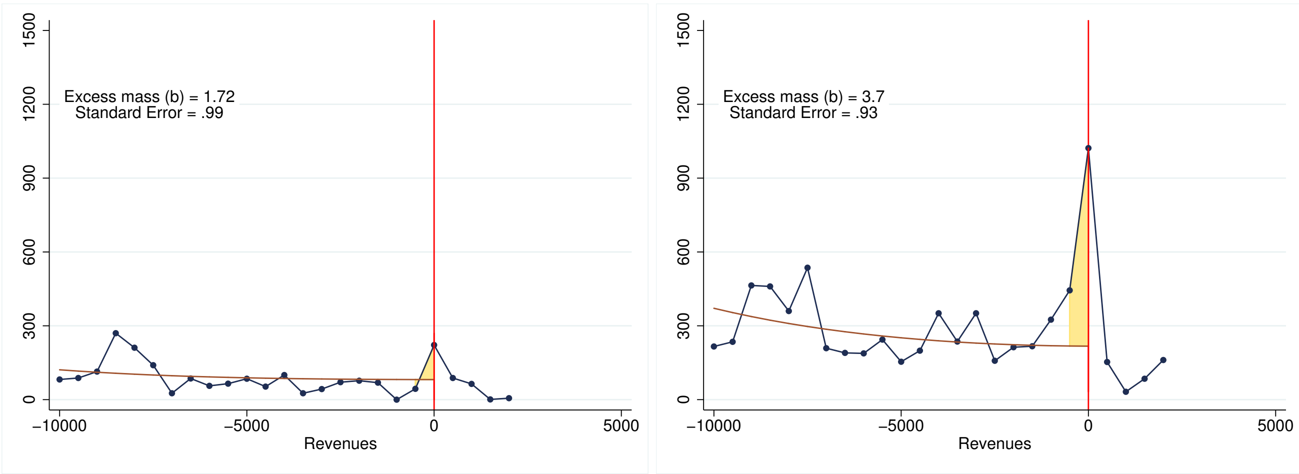
(b) 2002-2005



(c) 2006-2008

Figure 3.17: Bunching by Period

Notes: The figure shows bunching in different periods: 1999-2001, 2002-2005 and 2006-2008. Bunching increases over time until the introduction of the super-simplified regime in 2009. See the notes to Figure 3.6.



(a) Early Period 1999-2001 (b) Later Period 2002-2008

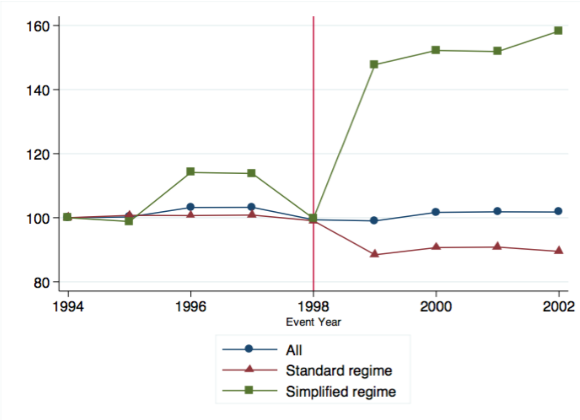
Figure 3.18: Increase in Income Shifting Within the Household

Notes: Panels (a) and (b) repeat panel (a) of Figure 3.16, splitting the period into two (an early period in panel (a) and a later period in panel (b)).

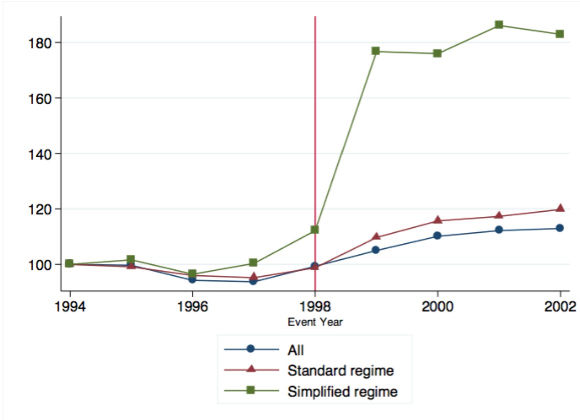
an almost equivalent decline in the number of agents in the standard regime.[32] The increase in mean income per business (panel (b)) was also instantaneous. For the simplified regime, this rise comes from the combination of an intensive and an extensive effect: entrepreneurs already in the simplified regime can remain in this regime while reporting much higher sales; and businesses in the standard regime in 1998 due to sales higher than the 100,000 Francs limit can switch to the simplified regime in 1999, to the extent they report a turnover below the new threshold. For the standard regime it mostly stems from this composition effect, as those switchers earn relatively less income compared with the rest of the firms remaining in the standard regime. All self-employed together, mean income per business barely increases.

Second and in contrast, the 2008 reform brings about complex changes through the introduction of a new fiscal regime. The number of self-employed in this new regime increases markedly for many years, this entry driven this time by newly self-employed individuals as the overall number of self-employed steadily increases (panel (c) in Figure 3.19). These new entrants had lower sales on average than entrepreneurs already in the simplified regime, despite the simultaneous increase in the threshold (from 27,000 to 32,000 €): the super-simplified regime was particularly tailored towards small businesses and indeed attracted smaller businesses year in year out (panel (d)).

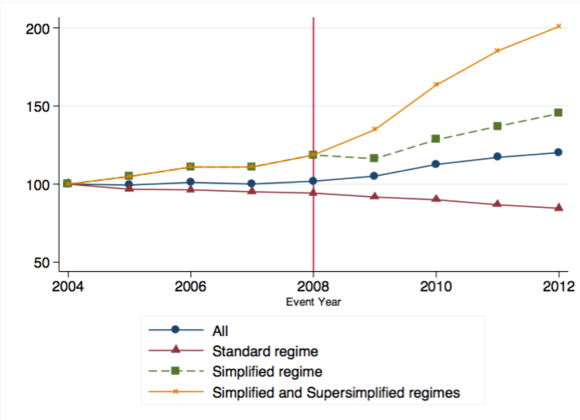
[32] Both series are normalized by their own 1994 value so the magnitudes on the graph do not allow to check that the increase in one is in fact almost equal to the decline in the other.



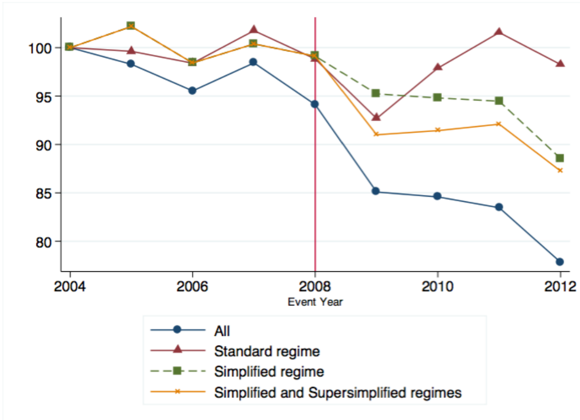
(a) Number of self-employed agents



(b) Average income per self-employed business



(c) Number of self-employed agents



(d) Average income per self-employed business

Figure 3.19: Event Studies: the 1999 and 2008 Reforms

Notes: We normalize all variables by their starting level in 1994 for panels (a) and (b) and in 2004 for panels (c) and (d). Panels (a) and (c) plot the number of self-employed agents in each of the standard and simplified regimes, as well as the total number of self-employed. Panels (b) and (d) plot the average income per self-employed business for those in the standard regime (net business income) and those in the simplified regime (gross revenues).

Third we consider smaller threshold's changes, which we see as examples of small and hardly noticeable adjustments. At 32,000 in 2009, the threshold increases to 32,100 in 2010, 32,600 in 2011-2013 and then to 32,900 in 2014-2015. Figure 3.20 shows these four thresholds with vertical lines, and plots the number of individuals reporting income in the simplified regimes from Industrial & Commercial Services or Non Commercial activities, by bins of 100 € and for each year between 2011 and 2015. Several lessons can be drawn. First and consistent with progressive learning, bunching at the threshold increases over time when the threshold does not change (from 2011 to 2013, and between 2014 and 2015). Second a mass of individuals is still visible at the 2013 threshold in 2014. This mass declines in 2015. Third and throughout all these years, a sizable mass of individuals report income just below 32,000 €, even though this threshold loses its legal value in 2010. Nothing else differentiates 32,000 from 31,000 – they both are as "round". The introduction of the super simplified regime and the accompanying extensive communication made the 32,000 € threshold highly salient. Subsequent changes in the eligibility threshold, which are much smaller, are far less visible. This persistent mass of individuals at the 2009 threshold is coherent with the costly learning hypothesis.

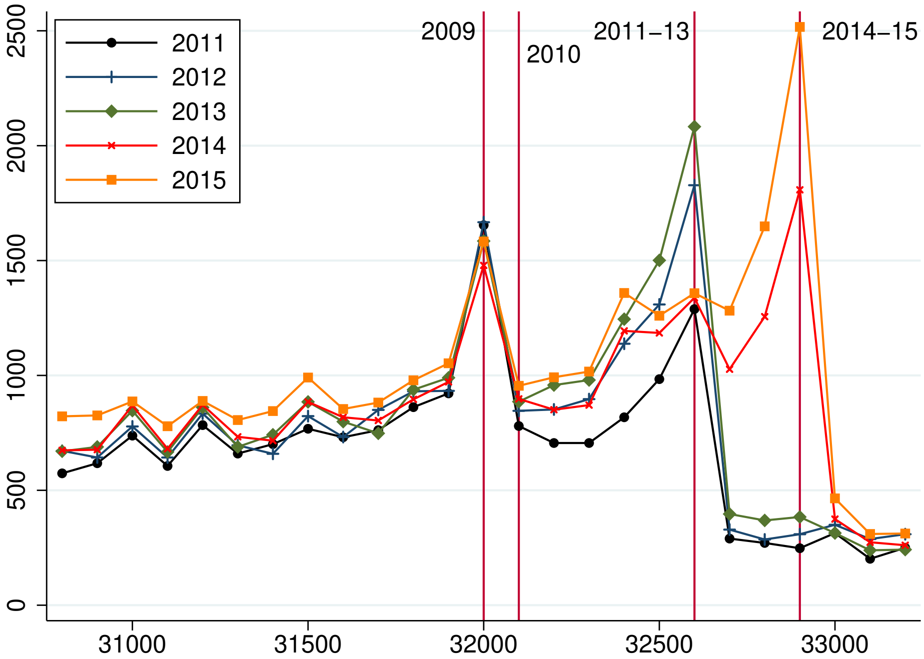


Figure 3.20: Bunching in each year 2011-2015

Notes: This Figure plots the number of individuals with income in the simplified regimes from the Industrial & Commercial Services and Non Commercial activities in 100 € bins around 32,000 €, for each year of the exhaustive dataset (2011-15). The vertical lines represent the eligibility thresholds over the years: 32,000 € (2009), 32,100 € (2010), 32,600 € (2011-13) and 32,900 € (2014-15).

3.5.5 Summarizing

All in all, we have highlighted different behavioral patterns close to the eligibility threshold, only in the regime and activities subject to this eligibility threshold, that cannot be well explained without tax evasion: individuals report more often round numbers, households with two self-employed in a simplified regime bunch more, and the bunching increases over time, which is consistent with individuals gradually understanding the mechanics of the tax code and how to play it. Saez (2010) and Chetty et al. (2013) also interpreted the bunching they observed for the self-employed as evasion.

3.6 The Value of Tax Simplicity

The previous section highlighted that tax evasion likely plays a key role in explaining the bunching observed below the thresholds of the simplified and super simplified regimes: individuals manage to remain in the simpler regimes to some extent by misreporting their revenues. However the motives behind the choice to report income below the threshold – whether this choice stems from a real activity response or tax evasion – remain unclear. Obviously monetary incentives are an essential factor: by reporting less income and/or filing for an often less taxed regime, individuals pay fewer taxes in the simpler regimes than they would in the standard one. But the sheer simplicity of these tax systems may also drive part of the bunching: filing under the simplified or super simplified regimes frees individuals from the administrative burdens associated with the standard regime. To disentangle monetary incentives from the simplicity motive and put a value on the tax simplicity, we develop a structural model and estimate its main parameters by fitting moments taken from the data.[33]

More precisely we jointly estimate the real income elasticities, the evasion elasticities and the value of tax simplicity, based on the bunching at the eligibility thresholds. The key to this estimation is first that the monetary gain from evasion scales with income while the gain from simplicity is fixed, and second that there is a large variability in the incentives faced by different types of agents at the same eligibility threshold because of their different tax brackets and the regimes they are in. Combined with the fact that these incentives have also changed extensively over time, even in periods where bunching has remained stable, overall we obtain a lot of different data moments that inform us about the parameters of interest. Ultimately, the structural parameters that can best explain the observed

[33] It is important to bear in mind that even if there is zero tax difference and zero tax base difference between the regimes, there is still an incentive to evade and choose a simple regime. First, the non zero tax rate is an incentive to underreport to avoid the taxes and social security contributions on the hidden sales. Second and by filing under a simple regime, an agent increases her chances to evade again in the future without being noticed (switching regimes might be perceived as a bad signal to the tax authority). And third in itself the tax simplicity has value. The least costly way is then often just to evade, rather than earn less income. Tax simplicity, monetary incentives and evasion are intrinsically related, which makes identification challenging. Part of the value of the tax simplicity of the simple regimes comes from the enhanced possibilities to evade taxes.

bunching across different tax brackets and years will turn out to display a large preference for tax simplicity, a sizable evasion elasticity, and a negligible real income elasticity.

3.6.1 Structural Estimation

Modelling the tax discontinuity

A given individual can operate in one of the three regimes described above: the simplified (regime "m") the super simplified (regime "f") and the standard one (regime "r"). Effective operating costs, taking into account input costs and VAT payments are modeled as a fraction c_i of revenues y_i in each of the regimes $i = m, f, r$. Each regime entails a "tax hassle cost" a_i, reflecting the tax reporting and compliance costs (e.g. administrative accounting requirements, costs of keeping track and complying with the tax procedure).

Let μ be the rebate on gross revenues in the simplified regime: the taxable income of agents in this regime is $(1-\mu)\cdot y$. The agent's effective average income tax rate (which is the same regardless of regime choice) is τ^y and the social insurance contributions rate in regime i is τ_i^{ss}. Denote by τ_i the effective rates in regime i, levied on the tax base applicable in that regime, denoted by z_i (which also differs across regimes). In practice, an agent's effective average income tax rate and his social insurance contribution rate depend on his total income (self-employed income, wages and salaries, ordinary capital income, etc.), household composition, activity type, and occupation, as explained in Section 3.2. We do take this heterogeneity into account in our numerical estimations when assigning a tax rate to each agent. For simplicity of the exposition in this section, we express the rates as if they were homogeneous across all agents in a given regime.

Agents in the simplified regimes can endogenously misreport their income. In accordance with the evidence provided in Sections 3.2 and 3.5, agents in the standard regime cannot easily misreport their income. Let the fraction of actual revenues which is reported be denoted by γ_i, where $\gamma_r = 1$ (no misreporting in the standard regime). Correspondingly, reported revenues in the simplified regime are $\hat{y}_m = \gamma_m y_m$ and in the super simplified regime $\hat{y}_f = \gamma_f y_f$. The effective rates and tax bases are as follows:

Standard regime:	$\tau_r = \tau^y + \tau_r^{ss}(1-\tau^y)$	is levied on net income	$z_r = (1-c_r)y_r$
Simplified regime:	$\tau_m = \tau^y + \tau_m^{ss}$	is levied on taxable (reported) income	$z_m = (1-\mu)\gamma_m y_m$
Super simplified regime:	τ_f	is levied on gross (reported) revenues	$z_f = \gamma_f y_f$

The notch created by the eligibility threshold generates a response in self-employed revenues, which we denote by Δy^*. This revenue response Δy^* can be estimated from the excess mass using the expression

$$B \approx f_0(y^*)\Delta y^*$$

where $f_0(y^*)$ is the counterfactual density at the threshold (i.e. the density that applies in the absence of the eligibility threshold).[34]

Structural model

We can map the observed bunching into structural elasticities and hassle costs, that requires specifying a parametric utility function. To infer the elasticities, we need to assume that agents who bunch come from a continuous interval of revenues above the eligibility threshold and that there is one agent who is the marginal buncher. We consider a static setting without optimization frictions.[35]

We parameterize the model as follows. Each agent has a type θ that captures his productivity and that reduces his cost of earning a given level of revenues. The disutility of generating revenues y for an agent of type θ is denoted by $h(y, \theta)$, increasing in y and decreasing in θ. We assume that:

$$h(y, \theta) = \frac{\theta}{1 + \frac{1}{\varepsilon}} \left(\frac{y}{\theta}\right)^{1+\frac{1}{\varepsilon}}$$

The cost of misreporting revenues in the simplified and super simplified regimes ($i = m, f$) is assumed to be isoelastic with elasticity η, namely:

$$g(y_i - \hat{y}_i) = \frac{\kappa}{1 + \frac{1}{\eta}} \left(\frac{y_i - \hat{y}_i}{\kappa}\right)^{1+\frac{1}{\eta}}$$

where κ is a scaling parameter that sets the right relation between the cost of real revenue generation $h()$ and evasion. In the standard regime, the misreporting cost is implicitly infinite, as it is hard to misreport when being part of a certified accounting center.

[34] As shown in Kleven and Waseem (2013) one can easily account for heterogeneous elasticities. For an agent with elasticity e, the earnings response is Δy_e^*. If $f_0(y, e)$ is the joint revenue-elasticity counterfactual distribution and $f_0(y) = \int_e f_0(y, e)$ is the unconditional counterfactual revenue distribution, the excess mass can be mapped to the average earnings response $\mathbb{E}(\Delta y_e^*) \approx f_0(y^*)E(\Delta y_e^*)$.

[35] Since we estimate elasticities assuming no optimization frictions, we are likely providing a lower bound of the elasticities. In these estimations, agents make a series of static decisions period by period. This is realistic in so far that for the small self-employed considered here, self-employed income is a flexible yearly choice variable, not subject to career concerns or strategic dynamic considerations, such as signaling.

An agent's utility from earning revenue y and reporting $\hat{y}$ in regime $i = m, f, r$ is thus:

$$u_i(y, \hat{y}) = y(1 - c_i) - T_i(\hat{y}) - h(y, \theta) - g(y - \hat{y}) - a_i$$

where $T_i(\hat{y})$ is the total tax liability as a function of reported revenues described above.

Simplified regime threshold: Let us start by describing what happens at the eligibility threshold for the simplified regime. Note that if there is no notch, in the simplified regime, the optimal choice of real and reported revenues of an agent are:

$$\begin{aligned} g'(y_m - \hat{y}_m) = \tau_m(1 - \mu) \quad \text{and} \quad & h'(y_m, \theta) = (1 - c_m) - g'(y_m - \hat{y}_m) \\ & \Rightarrow h'(y_m, \theta) = [(1 - c_m) - \tau_m(1 - \mu)] \end{aligned}$$

which implies that:

$$y_m = \theta[(1 - c_m) - \tau_m(1 - \mu)]^{\varepsilon} \tag{3.4}$$

$$\hat{y}_m = \theta[(1 - c_m) - \tau_m(1 - \mu)]^{\varepsilon} - \kappa[\tau_m(1 - \mu)]^{\eta} \tag{3.5}$$

With the eligibility threshold (notch), there is also a marginal agent $y^* + \Delta y^*$ who reports revenues exactly at the threshold y^* but would have reported revenues at $y^* + \Delta y^*$ absent the notch. If he were unconstrained by the notch, his choice would be characterized by reported revenues:

$$y^* + \Delta y^* = (\theta^* + \Delta\theta^*)[(1 - c_m) - \tau_m(1 - \mu)]^{\varepsilon} - \kappa[\tau_m(1 - \mu)]^{\eta} \tag{3.6}$$

and actual revenues:

$$y_m = (\theta^* + \Delta\theta^*)[(1 - c_m) - \tau_m(1 - \mu)]^{\varepsilon}$$

With the notch, this agent reports revenues at the notch, but his actual revenues are $y_m^* = y_m(y^*)$ as a function of the report, where y_m^* is given by:

$$(1 - c_m) - h'(y_m^*, \theta^* + \Delta\theta^*) - g'(y_m^* - y^*) = 0 \Rightarrow 1 - c_m - \left(\frac{y_m^*}{\theta^* + \Delta\theta^*}\right)^{\frac{1}{\varepsilon}} - \left(\frac{y_m^* - y^*}{\kappa}\right)^{\frac{1}{\eta}} = 0 \tag{3.7}$$

This agent's utility at the notch is thus:

$$u_m^* = y_m^*(1 - c_m) - \tau_m(1 - \mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m$$

Let's denote by y_r^I the indifference point in the standard regime, such that the agent is indifferent between earning revenues y_m^* and reporting revenues exactly equal to the threshold y^* or earning y_r^I (which is actual revenues, since there is no misreporting in the standard regime). y_r^I is interior, and hence characterized by the optimality (tangency) condition in standard regime:

$$y_r^I = (\theta^* + \Delta\theta^*)[(1 - c_r)(1 - \tau_r)]^{\varepsilon} \tag{3.8}$$

The indifference condition is expressed as:

$$\begin{aligned} y_r^I(1 - c_r)(1 - \tau_r) - h(y_r^I, \theta^* + \Delta\theta^*) - a_r = \\ y_m^*(1 - c_m) - \tau_m(1 - \mu)y^* - h(y_m^*, \theta^* + \Delta\theta^*) - g(y_m^* - y^*) - a_m \end{aligned} \tag{3.9}$$

For given primitives $(\varepsilon, \eta, c_r, c_m, a_r, a_m, \kappa)$ and policy parameters $(y^*, \tau_m, \tau_r, \mu)$, we can solve for the variables $(y_m^*, \Delta y^*, y_r^I, \theta^* + \Delta\theta^*)$ using the system of nonlinear equations (3.6)-(3.9).

Super simplified regime threshold: For the super simplified regime threshold, everything is exactly the same as for the simplified regime, but using the appropriate tax rates. Equations (3.6)-(3.9) still apply, replacing c_m with c_f, y_m with y_f, y_f^* with y_f^* and $\tau_m(1 - \mu)$ with τ_f. In this case, for given primitives $(\varepsilon, \eta, c_r, c_f, a_r, a_f, \kappa)$ and policy parameters $(y^*, \tau_f, \tau_r, \mu)$, we can solve for $(y_f^*, \Delta y^*, y_r^I, \theta^* + \Delta\theta^*)$ as well.

Ultimately, the solution of this system of equations yields a model predicted $\Delta y^*(\varepsilon, \eta, c_r, c_f, a_r, a_f, \kappa;)$ as a function of the primitives and policy parameters.

Intuitively, a given observed bunching can be generated by i) a preference for low tax hassle, ii) a direct response to tax incentives (embodied in the real elasticity ε), and iii) an evasion response (embodied in the misreporting elasticity η). The hassle cost by itself creates a "pure notch" – i.e. an increase in the average tax without a change in the marginal tax; even in the absence of monetary incentives, tax accounting and reporting requirements act like an average tax increase and have distortionary effects.

Structural Estimation Method

We now explain how we structurally estimate the model using a generalized method of moments. Different agents in different years face widely different incentives because they are in different tax brackets (recall the peculiarity of the French tax system as described in Section 3.2), in different regimes, and because income taxes, social security contribution rates and rebates have changed over time. As a result, we have many data moments, i.e. observations of the empirical responses Δy^* across tax brackets, regimes and years, which we can target in order to find the parameters that fit best. Note that we will only use the recent period 2011-2015 for which we have full population data and during which, as discussed, agents' responses stabilize. And we focus attention on the I&C Services, because individuals at a given income level in Non-Commercial (NC) activities may face a wide heterogeneity in their social contributions, which are not observable in our data.

Formally, let i index the tax bracket, n index the regime (super simplified or simplified) and t index groups of years for which the thresholds y^* and rebates μ are constant. For each regime, activity, tax bracket, and year, there is a model-predicted bunching interval Δy^*_{nit}. Its empirical counterpart in the data is $\hat{\Delta} y^*_{nit}$. Each $\hat{\Delta} y^*_{nit}$ is thus a data moment.

The parameters we seek to estimate are the operating costs, hassle costs, real income elasticities and evasion elasticities for the various regimes and tax brackets. In principle, there can be heterogeneous parameters for each regime, activity, year, and tax bracket. We make the following assumptions, which allow us to limit the number of parameters estimated: (i) the tax hassle costs and the operating costs are allowed to differ by regimes, but are the same across tax brackets; (ii) the elasticities are allowed to differ by tax brackets but are the same across regimes. As a result, we have a vector χ of 13 parameters to be estimated with:

$$\chi := ((\eta_i)_{0\leq i\leq 3}, (\varepsilon_i)_{0\leq i\leq 3}, (a_n)_{n\in\{m,f\}}, (c_n)_{n\in\{m,f\}})$$

where η_i is the evasion elasticity in tax bracket i, ε_i is the real income elasticity in tax bracket i, a_n and c_n are the hassle and operating costs in the I&C Services activities in regime n.

We have $M = 16$ moments (four tax brackets times two regimes times two sets of years during which the policy parameters were constant). The loss function we minimize is $L(\chi_n)$:

$$L(\chi_n) = \sum_{m=1}^{M} \frac{1}{M} \left(\hat{\Delta} y^*_{nkit} - \Delta y^*_{nkit} \right)^2 \tag{3.10}$$

3.6.2 Estimation Results

Table 3.7 shows the estimation results. Three main results stand out. First the real income elasticities ε_i are very small (less than 1%). Second, the evasion elasticities are high and increase with the tax bracket.

Third, hassle costs are estimated much lower in the super simplified regime than in the simplified regime. Importantly, it is not possible to match the data moments correctly without tax simplicity (by imposing the hassle costs to be 0), especially for the super simplified regime. The appeal of tax simplicity thus seems much stronger in the super simplified regime. Finally, operating costs for the simplified and super-simplified regimes are estimated at around 50% of sales, which turns out to be the rebate chosen by the tax authority.

To conclude this analysis, it might be helpful to compare our results with other findings in the literature or with independent computations by French tax authorities. First, the Inspection Generale des Finances (IGF) finds an average estimate of 400 euro of tax evasion per self-employed individual (i.e averaging between individuals in the simplified and super-simplified regimes) based on direct ex

Table 3.7: Structural Estimation Results

Parameter		Description	Value
ε		Frisch elasticity in …	
	ε_0	… tax bracket 0	< 0.01
	ε_1	… tax bracket 1	< 0.01
	ε_2	… tax bracket 2	< 0.01
	ε_3	… tax bracket 3	< 0.01
η		Cheating elasticity in …	
	η_1	… tax bracket 1	0.21
	η_1	… tax bracket 1	0.76
	η_2	… tax bracket 2	0.79
	η_3	… tax bracket 3	0.79
Δa		Hassle costs difference in euros between …	
	$a_r - a_m$	… real and simplified regimes	85
	$a_r - a_f$	… real and super simplified regimes	524
c		Operating costs as a fraction of revenues in…	
	c_m	… simplified regime	0.44
	c_f	… super simplified regime	0.51
κ		Scale parameter in the cheating function	872

Notes: The table shows the results from the full structural estimation described in Section 3.6. The period is 2011-2015 for both simplified and super simplified regimes. See Section 3.6 for the computational details. Hassle costs are expressed in euros.

post tax auditing. Second, Pitt and Slemrod (1989) find that allowing for individual itemization entails a cost equal to 0.12% of total gross revenues; Benzarti (2017) finds an itemization cost of about 0.7% of total gross revenues. Benzarti then shows that this cost amounts to around 15 working hours per year. 400 € correspond to 27 hours of work at the hourly mean net wage of 14.6 € in 2015.

3.7 Conclusion

We study how French self-employed respond to a notch in their tax schedule. We show that they bunch massively below the eligibility threshold of the simplified and super simplified regimes above which they no longer qualify for these regimes. They also clearly respond to tax incentives as agents with higher marginal tax rates, who gain more financially from filing under a simpler regime, bunch more than other individuals. Even though it is not possible from the data to directly separate tax evasion from a real activity response, we provide three pieces of evidence suggesting that at least some of this bunching comes from tax evasion. First, the tax returns are more often round numbers close to the threshold than far from it, an indication that the reported figure is more likely to be forged. Second, in households with two self-employed individuals, the highest earner appears to shift some of her income to her partner as she approaches the threshold. Third bunching increases over time, and agents react rapidly to a large and salient institutional change whereas they respond slowly to a small or complex one.[36]

The bunching reveals the entrepreneurs' willingness to file under the simpler regimes – possibly misreporting their income to achieve that goal – but not their incentives to do so: individuals may be guided by pure monetary incentives, or they may be attracted by the tax simplicity of these regimes. To disentangle between the two motives, we estimate a structural model by fitting data moments. Our estimation confirms the role of tax evasion and tax simplicity in explaining the observed bunching. The simplicity motive appears particularly important in the super simplified regime, where it is valued at around 500 € (versus less than 100 € in the simplified regime).

We do not propose a welfare analysis of the introduction of the simpler regimes and in particular cannot conclude that public finances deteriorate with them. These regimes were explicitly introduced to facilitate the creation of firms that would otherwise not exist, and to shift work from the informal to the formal sector. Consistent with this, Barruel et al. (2012) shows that three quarters of the firms created in the first semester of 2010 under the super simplified regime would not have been created without the introduction of this regime; and we observe after 2009 a large number of creations of super simplified firms, without a meaningful decrease in the number of firms in the standard or simplified regimes. So the public finances losses induced by tax evasion, which happen at the margin of the

[36] It would therefore be illusory in most cases to use sales from the simplified or super simplified regimes as a RDD instrument, as this variable is to some extent manipulated around the threshold.

thresholds, may turn out small compared to the overall gains. While this assessment remains beyond the scope of our analysis, it would be a promising avenue for future research to evaluate the general equilibrium effects of the existence of the simplified and super simplified regimes and their impact on public finances and welfare.

Our analysis could be extended in several other interesting directions. A second avenue for future research would be to study whether tax simplicity improves the chances of success of a self-employed activity: do the self-employed individuals who understand tax incentives better end up doing better even in the long-run? Do they become true "entrepreneurs" and ultimately job creators? A third avenue would be to reflect on policies that can reduce the regressivity caused by the complexity of the tax system, such as the provision of tax workshops or tax accounting services to the lower educated or lower income agents who appear to be making worse financial choices. A fourth avenue would be to move to panel data and analyze how self-employment opportunities affect individuals' income mobility and also their ability to accumulate or preserve human capital over their lifetime. These and other extensions await further research.

Bibliography

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "House prices, collateral, and self-employment," *Journal of Financial Economics*, 2015, *117* (2), 288–306.

Barruel, Frédéric, Stéphane Thomas, Serge Darriné, and Henri Mariotte, "Trois auto-entrepreneurs sur quatre n'auraient pas créé d'entreprise sans ce régime," *Insee Première*, 2012, *1388.*

Berson, Clémence, Raphaël Lardeux, and Claire Lelarge, "Wage volatility, as perceived by workers," *In progress*, 2019.

Best, Michael and Henrik Kleven, "Housing Market Responses to Transaction Taxes: Evidence From Notches and Stimulus in the UK," *Working paper*, 2016.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus revenue efficiency with limited tax capacity: theory and evidence from Pakistan," *Journal of Political Economy*, 2015, *123* (6), 1311–1355.

Best, Michael, James Cloyne, Ethan Ilzetzki, and Henrik Kleven, "Interest rates, debt and intertemporal allocation: evidence from notched mortgage contracts in the United Kingdom," *Working paper*, 2015.

Bruce, Donald, "Effects of the United States tax system on transitions into self-employment," *Labour Economics*, 2000, *7*, 545–574.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 2017, *9* (2), 144–164.

Chetty, Raj, John N Friedman, and Emmanuel Saez, "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *The American Economic Review*, 2013, *103* (7), 2683–2721.

_ , John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records.," *Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

Cour des Comptes, "Les organismes de gestion agréés, 40 ans après," Technical Report 2014.

Cullen, Julie and Roger Gordon, "Tax Reform and Entrepreneurial Activity," *Tax Policy and the Economy*, 2006, *20*, 41–72.

_ **and** _ , “Taxes and Entrepreneurial Risk Taking: Theory and Evidence for the U.S,” *Journal of Public Economics*, 2007, *91* (7-8), 1479–1505.

Eissa, Nada and Hilary Hoynes, “Taxes and the labor market participation of married couples: the earned income tax credit,” *Journal of Public Economics*, 2004, *88*, 1931–1958.

_ **and** _ , “The Hours of Work Response of Married Couples: Taxes and the Earned Income Tax Credit,” in Jonas Agell and Peter Birch Sorensen, eds., *Tax Policy and Labor Market Performance*, MIT Press, 2006.

Fack, Gabrielle and Camille Landais, “Are Tax Incentives For Charitable Giving Efficient? Evidence from France,” *American Economic Journal: Economic Policy,*, 2010, *2* (2), 117–141.

_ **and** _ , “The Effect of Tax Enforcement on Tax Elasticities: Evidence from Charitable Contributions in France,” *Journal of Public Economics*, 2016, *1330* (1), 23–40.

Feldman, Naomi and Joel Slemrod, “Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns,” *Economic Journal*, 2007, *17* (518), 327–352.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty, “Inequality Dynamics in France, 1900-2014: Evidence from Distributional National Accounts (DINA),” 2016.

_ **,** _ **, and** _ , “Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014),” 2017.

Gelber, Alex, “Taxation and the Earnings of Husbands and Wives: Evidence from Sweden,” *Review of Economics and Statistics*, 2014, *96* (2), 287–305.

Gelber, Alexander M, Damon Jones, and Daniel W Sacks, “Estimating Earnings Adjustment Frictions: Method and Evidence from the Social Security Earnings Test,” Technical Report 2017.

_ **,** _ **,** _ **, and Jae Song**, “Using Kinked Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test,” Technical Report 2017.

Gentry, William M. and R. Glenn Hubbard, “Tax Policy and Entrepreneurial Entry,” *The American Economic Review*, 2000, *90* (2), 283–287.

Gordon, Roger H. and Joel Slemrod, “Are ’Real’ Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases?,” in J. Slemrod, ed., *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, Russell Sage Foundation and Harvard University Press, 2000, pp. 240–280.

Gruber, Jonathan and Emmanuel Saez, “The elasticity of taxable income: evidence and implications,” *Journal of Public Economics, Elsevier*, 2002, *84* (1), 1–32.

Hamilton, Barton H., "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, 2000, *108* (3), 604–631.

Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar, "Can Unemployment Insurance Spur Entrepreneurial Activity?," *National Bureau of Economic Research Working Paper*, 2017, (20717).

Hoopes, Jeffrey, Joel Slemrod, and Daniel Reck, "Taxpayer Search for Information: Implications for Rational Attention," *American Economic Journal: Economic Policy, forthcoming*, 2017.

Katz, Lawrence and Alan Krueger, "The Role of Unemployment in the Rise in Alternative Work Arrangements," *American Economic Review Papers and Proceedings*, 2017, *107* (5), 388–392.

Katz, Lawrence F. and Alan B. Krueger, "The Rise and Nature of Alternative Work Arrangements in the United States, 1995-2015," *NBER Working Paper*, 2016, *22667.*

Kleven, Henrik and Mazhar Waseem, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan.," *Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Kobak, Dmitry, Sergey Shpilkin, and Maxim S. Pshenichnikov, "Integer Percentages as Electoral Falsification Fingerprints," *The Annals of Applied Statistics*, 2016, *10* (1), 54–73.

Landais, Camille and Johannes Spinnewijn, "Unemployment, Consumption Smoothing And the Value of Insurance," *London School of Economics Working paper*, 2017.

_ , Jonas Kolsrud, Peter Nilsson, and Johannes Spinnewijn, "The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden," *IZA Discussion Paper*, 2015, (9185).

Lelarge, Claire, David Sraer, and David Thesmar, "Entrepreneurship and Credit Constraints: Evidence From a French Loan Guarantee Program," in Josh Lerner and Antoinette Schoar, eds., *Differences in Entrepreneurship*, 2008.

Lockwood, Benjamin B., "Optimal Income Taxation with Present Bias," *University of Pennsylvania Working Paper*, 2016.

_ and Dmitry Taubinsky, "Regressive Sin Taxes," *University of Pennsylvania Working Paper*, 2016.

Pomeranz, Dina, "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–2569.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points ?," *American Economic Journal : Economic Policy*, 2010, *2* (3), 180–212.

__ , **Joel Slemrod, and Seth Giertz**, "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 2012, *50* (1), 3–50.

Schmalz, Martin C, David A Sraer, and David Thesmar, "Housing collateral and entrepreneurship," *The Journal of Finance*, 2016.

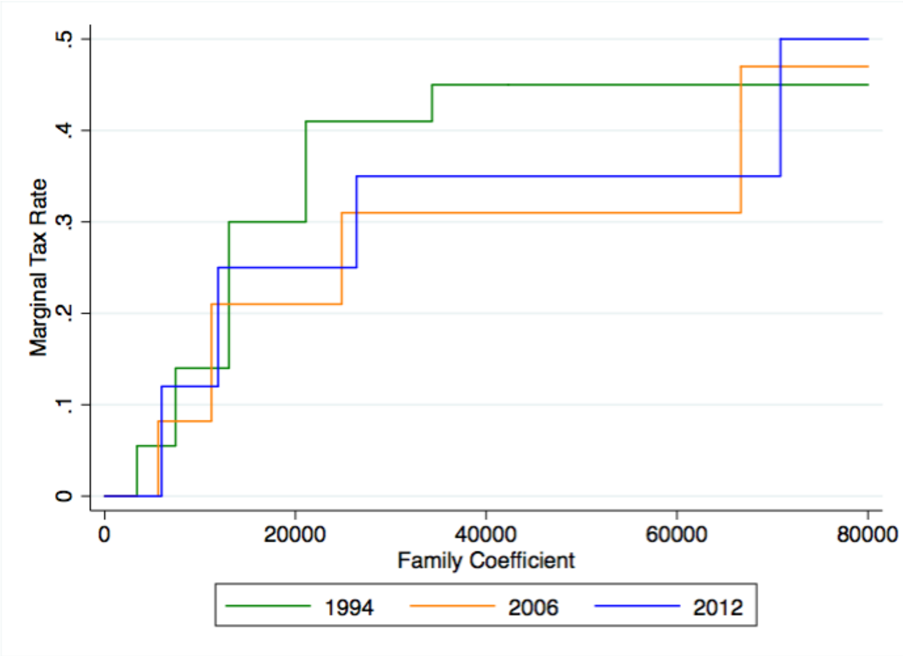
Schoar, Antoinette, "The divide between subsistence and transformational entrepreneurship," *Innovation Policy and the Economy*, 2010, *10* (1), 57–81.

Slemrod, Joel and James Salleel, "Car Notches: Strategic Automaker Responses to Fuel Economy Policy," *Journal of Public Economics*, 2012, *96* (11-12), 981–999.

APPENDIX

3.A Additional Figures

Figure 3.21: Income Tax Rates in France 1994-2015



Notes: The figure shows the marginal tax rates in different tax brackets and for a selected set of years. The x axis shows the family coefficient, i.e. taxable income divided by the number of parts, as explained in Section 3.2.

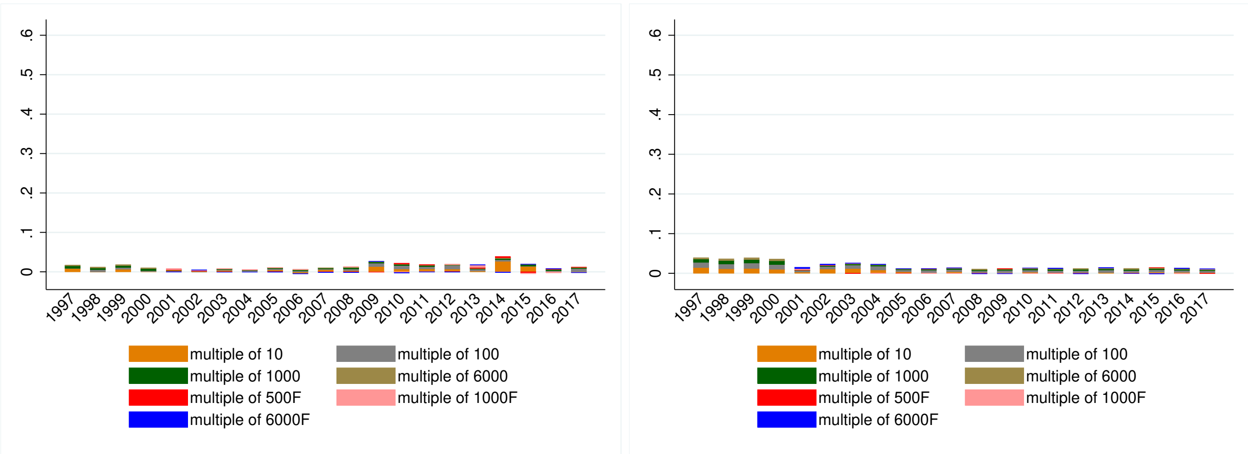


Figure 3.22: Excess share of declarations in the standard regime (left) and for wages (right), by round number

3.B Additional Tables

Table 3.8: Summary Statistics: Self-Employed and Wage Earners 1994-1998

	All	**With wage income only**	**With self-employed income only**	**With any self-employed income**
Panel A				
Age	40	39	48	47
Married or in Civil Union	0.56	0.55	0.68	0.68
Has any Children	0.43	0.43	0.42	0.44
Number of children	0.77	0.77	0.76	0.79
Retired	0.08	0.07	0.16	0.14
Panel B				
Wage Income	18377	19329	0.00	5275
Self-employed Income	1927	0	30254	28438
Capital Income	1457	1250	3642	4306
Panel C				
Zero Tax bracket	0.19	0.20	0.16	0.14
Low Tax bracket	0.36	0.37	0.26	0.25
Medium Tax bracket	0.31	0.31	0.28	0.28
High Tax bracket	0.14	0.12	0.30	0.32
Panel D				
Population (in mill.)	130	121.2	6.8	8.8

Notes: See Table 3.2.

Table 3.9: Summary Statistics: Self-Employed and Wage Earners 1999-2008

	All	**With wage income only**	**With self-employed income only**	**With any self-employed income**
Panel A				
Age	40	40	49	48
Female	0.46	0.47	0.31	0.31
Married or in Civil Union	0.50	0.49	0.64	0.63
Has any children	0.40	0.40	0.39	0.40
Number of Children	0.70	0.70	0.70	0.72
Retired	0.05	0.04	0.16	0.14
Claimed unemployment benefits	0.09	0.10	0.02	0.03
Panel B				
Wage Income	19750	20711	0	5824
Self-employment Income	2066	0	34963	32032
Capital Income	2203	1904	5531	6533
Panel C				
Zero Tax bracket	0.15	0.15	0.14	0.13
Low Tax bracket	0.32	0.32	0.22	0.21
Medium Tax bracket	0.38	0.38	0.30	0.31
High Tax bracket	0.15	0.14	0.33	0.35
Panel D				
Population (in mill.)	283.8	265.5	13.7	18.3

Notes: See Table 3.2.

Table 3.10: Summary Statistics: Self-Employed and Wage Earners 2009-2012

	All	**With wage income only**	**With self-employed income only**	**With any self-employed income**
Panel A				
Age	41	41	50	48
Female	0.48	0.49	0.35	0.36
Married or in Civil Union	0.45	0.44	0.57	0.55
Has any Children	0.39	0.39	0.36	0.37
Number of children	0.67	0.67	0.63	0.66
Retired	0.07	0.06	0.21	0.17
Claimed unemployment benefits	0.13	0.14	0.04	0.07
Panel B				
Wage Income	20470	21503	0	7145
Self-employment Income	1939	0	31506	26980
Capital Income	2803	2493	6008	6801
Panel C				
Zero Tax bracket	0.13	0.13	0.18	0.16
Low Tax bracket	0.29	0.30	0.20	0.20
Medium Tax bracket	0.46	0.47	0.36	0.38
High Tax bracket	0.11	0.10	0.26	0.26
Panel D				
Population (in mill.)	118.8	110.3	5.8	8.5

Notes: See Table 3.2.

Table 3.11: Full Structural Estimation Results– Tax Hassle Costs and Elasticities

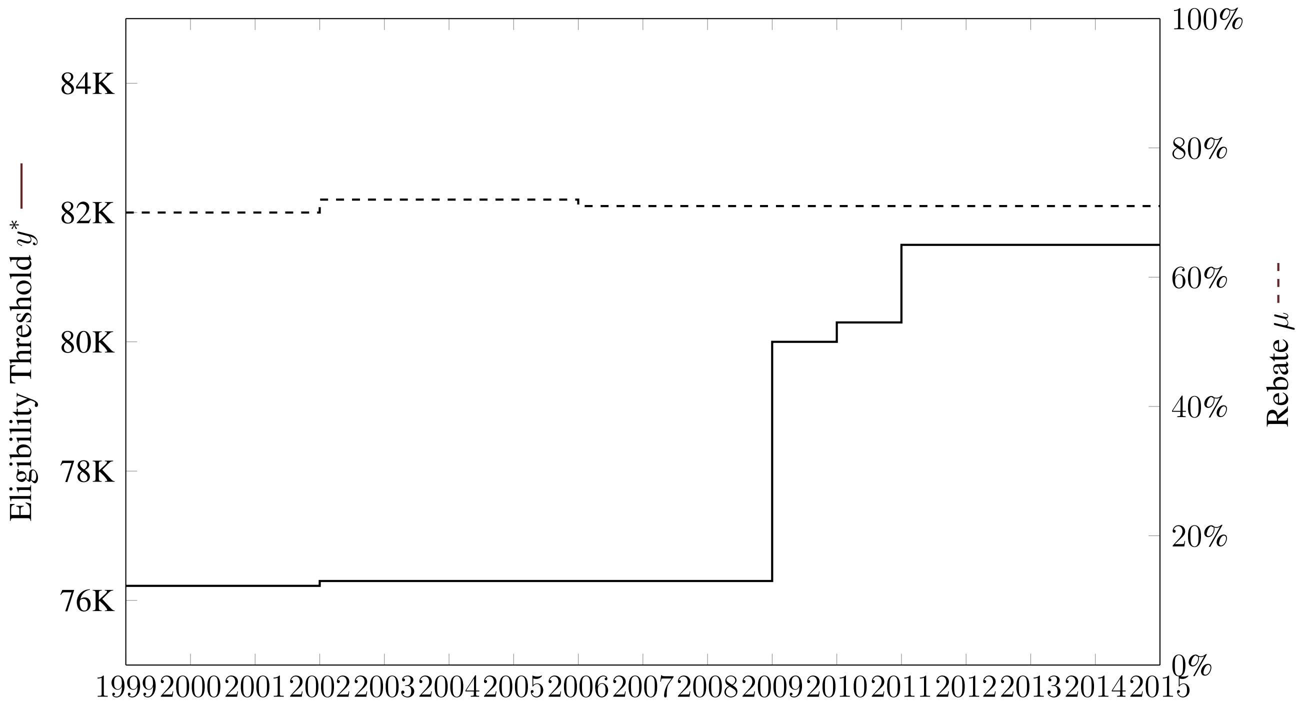
Cost I&C Services (% of rebate)	Cost Non Commercial (% of rebate)	Hassle Cost I&C Services a_S	Hassle Cost Non Commercial a_{NC}	Tax bracket	Structural Elasticity e
Panel A – Simplified Regime					
				1	0.01
0.5	0.1	315	456	2	0.02
				3	0.06
Panel B – Super Simplified Regime					
0.3	0.3	162	648	1	0.08
				2-3	0.01
Panel C – By Additional Income Sources					
With salaried income					
				1	0.01
0.5	0.2	304	145	2	0.03
				3	0.07
Without salaried income					
				1	0.02
0.5	0.2	149	144	2	0.01
				3	0.04
With pension income					
0.5	0.2	305	580	1-2-3	0.02
Without pension income					
0.5	0.2	150	299	1-2-3	0.01

Notes: The table shows the results from the full structural estimation described in Section 3.6. For the simplified regime, the period is 2006-2012. For the super simplified, it is 2009-2012. See Section 3.6 for the computational details. Hassle costs are expressed in euros.

3.C Online Appendix

3.C.1 Additional Tables and Figures

Figure 3.23: Eligibility Thresholds and Rebate for the Simpler Regimes (Retail)



Notes: The figures shows the eligibility threshold in euros along with the rebates for I&C retail activities. Before 1999, the I&C retail is not distinguished from the I&C services.

3.C.2 French Tax Calculation Primer

The French tax schedule typically looks as follows:

Bracket	Lower Bond	Upper bond	Marginal rate
1	$\underline{y_0} = 0$	$\underline{y_1}$	τ_1
2	$\underline{y_1}$	$\underline{y_2}$	τ_2
3	$\underline{y_2}$	$\underline{y_3}$	τ_3
4	$\underline{y_3}$	$\underline{y_4}$	τ_4
5	$\underline{y_4}$	∞	τ_5

In order to determine the tax amount to be paid by a household, the first thing to compute is the Family coefficient y which is defined as the ratio between taxable income Y and the number of parts N of the household :

$$y = \frac{Y}{N} \tag{3.11}$$

The household that has a family coefficient $y \in [\underline{y_{M-1}}; \underline{y_M}]$ belongs to the bracket M. Then, the amount of tax the household has to pay is :

$$T(y,N) = N \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y - \underline{y_{M-1}})] \tag{3.12}$$

For instance, for a household with a family coefficient $y \in [\underline{y_2}; \underline{y_3}]$, we have :

$$T(y,N) = N \times (\tau_1 \times \underline{y_1} + \tau_2 \times (\underline{y_2} - \underline{y_1}) + \tau_3 \times (y - \underline{y_2})) \tag{3.13}$$

Cap of the Family Coefficient

Let's assume that the number of parts is $N_b + N_a$ where N_b is the base number of parts, and N_a is the additional number of parts. To calculate the cap, one first calculates the tax that would apply without the additional parts: $y^b = Y/N_b$. We must then consider two possible situations : if the additional number of parts N_a (i) does place the household in a higher tax bracket, or (ii) does not place the household in a higher tax bracket.

Situation 1

If the additional number of parts N_a does not place the household in a higher tax bracket, then :

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (y^b - \underline{y_{M-1}})] \tag{3.14}$$

The difference in taxes is :

$$T(y^b, N_b) - T(y,N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b y^b - N_b \underline{y_{M-1}} - Ny + N\underline{y_{M-1}}) \tag{3.15}$$

By definition, we have $Y = N_b y^b = Ny$, then :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times \underline{y_{M-1}}(N - N_b) \tag{3.16}$$

We can re-arrange the expression to obtain :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times [\sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) - \tau_M \times \underline{y_{M-1}})] \tag{3.17}$$

Situation 2

If the additional number of parts N_a places the household in a higher tax bracket, then :

$$T(y^b, N_b) = N_b \times [\sum_{m=1}^{M} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_{M+1} \times (y^b - \underline{y_M})] \tag{3.18}$$

The difference in taxes is :

$$T(y^b, N_b) - T(y, N) = (N_b - N) \times \sum_{m=1}^{M-1} \tau_m \times (\underline{y_m} - \underline{y_{m-1}}) + \tau_M \times (N_b \underline{y_M} - N_b \underline{y_{M-1}} - Ny + N\underline{y_{M-1}}) + \tau_{M+1} N_b \times (y^b - \tag{3.19}$$

Chapter 4

Semi-Structural VAR and Unobserved Components Models to Estimate Finance-Neutral Output Gap

Semi-Structural VAR and Unobserved Components Models to Estimate Finance-Neutral Output Gap

GÁBOR KÁTAY LISA KERDELHUÉ

MATTHIEU LEQUIEN

Addresses - Gábor Kátay : European Commission, Joint Research Centre (JRC), Lisa Kerdelhué : Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE and Banque de France, Matthieu Lequien :Banque de France, Insee and PSE.
We owe a special thank to Michel Juillard, Paolo Paruolo and Phurichai Rungcharoenkitkul for their careful reading of our manuscript and their many valuable comments and suggestions. We are grateful to Máté Tóth for sharing with us the codes of the semi-structural UCM in Melolinna and Tóth (2019) and the useful inputs he provided. We also thank the participants of the seminar at the Banque de France and at the European Central Bank, as well as the members of the Working Group on Econometric Modelling of the European System of Central Banks for the valuable ideas they raised in discussion. Matthieu Lequien thanks the support of the EUR grant ANR-17-EURE-0001. The views expressed in this paper are those of the authors and not necessarily those of the institutions the authors are affiliated with.

Abstract

The paper assesses the impact of adding information on financial cycles on the output gap estimates for eight advanced economies using two unobserved components models: a reduced form extended Hodrick-Prescott filter, and a standard semi-structural unobserved components model. To complement these models, a semi-structural vector autoregression model is proposed in which only supply shocks are identified. The accuracy of the output gap estimates is assessed based on their performance in predicting recessions. The models with financial variables generally produce more accurate output gap estimates at the expense of increased real-time volatility. While the extended Hodrick-Prescott filter is particularly appealing for its real-time stability, it lags behind the two semi-structural models in terms of forecasting performance. The vector autoregression model augmented with financial variables features the best in-sample forecasting performance, and it has similar real-time prediction capabilities to the semi-structural unobserved components model. Overall, financial cycles appear to be relevant in Japan, Spain, the UK, and – to a lesser extent – in the US and in France, while they are relatively muted in Canada, Germany, and Italy.

Keywords: unobserved components model; semi-structural VAR; output gap; financial cycle; sustainable growth; credit; house prices; advanced economies

JEL classification: C32; E32; E44; G01; O11; O16

4.1 Introduction

Potential output has been traditionally defined as the maximum level of economic activity attainable without triggering inflation and, in the same context, as the output linked to the level of employment that results in a nonaccelerating rate of inflation (Okun (1962)). The observed empirical regularity between output fluctuation and the cyclical pattern of inflation or unemployment has been for a long time the key ingredients in a large variety of statistical filters, reduced form and general equilibrium models aiming to estimate the potential growth and the output gap. In such models, financial factors are either completely ignored or when included, their roles are limited to amplify the persistence of the shocks by slowing down somewhat the return of the economy to its steady state path (see e.g. Kiyotaki and Moore (1997); Bernanke et al. (1999); Woodford (2003)). However, historical evidence suggests that unsustainable developments in the financial and housing sectors can generate large imbalances in the economy even if inflation and unemployment are low and stable (see e.g. White (2006); Hume and Sentance (2009); Schularick and Taylor (2012); Jordà et al. (2013)).

The present paper is closely related to the recent empirical literature that extends traditional statistical output gap estimation techniques by incorporating financial cycle information. We start by conceptually and empirically comparing the performance of two types of popular unobserved components models (UCMs) in extracting both demand-driven traditional business cycles and financial cycles. The first UCM follows a novel "reduced form" approach pioneered by Borio et al. (2017) to estimate the "finance-neutral" output gap. In the "extended Hodrick-Prescott (HP)" model, pre-transformed financial cycle indicators are directly incorporated as additional covariates into the state-space representation of the univariate HP filter (model A)[1]. The second UCM is a simple semi-structural model featuring a Phillips curve, an Okun's law, a stochastic process relating output gap to capacity utilisation, and a separate block that relates financial cycles to the output gap (model B). Both the reduced form and the semi-structural UCMs augmented with financial variables are gaining increasing popularity[2], however, their relative advantages and limitations have not yet been empirically assessed.

[1]We also estimate a more flexible representation of this model, allowing for serial correlation of the output gap and introducing shocks to the level of the trend. However, the results from these two versions are very similar in all aspects. In other words, the additional flexibility of the dynamic multivariate filter compared to extended HP has very limited practical relevance. The results from this alternative specification are reported in Appendix 4.B.

[2]Following Borio et al. (2017), several subsequent papers adopt the same (or very similar) methodology to recover the "sustainable growth" and the "finance-neutral output gap" (see e.g. Anvari et al. (2014); Odor and Kucserova (2014); Felipe et al. (2015); Krupkina et al. (2015); Maliszewski and Zhang (2015); Berger et al. (2015); Alberola-Ila et al. (2016); Amador-Torres et al. (2016); Grintzalis et al. (2017)). Bernhofer et al. (2014) adopt a more general statistical filtering technique proposed by Harvey and Jaeger (1993) to estimate the finance neutral output gap for several advanced and emerging EU countries. However, the basic concept of the method remains the same: financial cycle indicators are directly incorporated into output gap equation to help to better explain the cyclical movements of the output. Other papers extend the previous approach by incorporating financial information into semi-structural UCMs (see e.g. Rünstler and Vlekke (2018); Melolinna and Tóth (2019)). Our baseline semi-structural model is mostly based on Melolinna and Tóth (2019).

To complement these previous models, we propose a new "semi-structural vector autoregression (VAR)" model with long-run restrictions (model C). The model is a modified version of Blanchard and Quah (1989) in which we exploit a wider set of information by using several (business and financial) cycle indicators without imposing further restrictions. It extends the original Blanchard-Quah model by combining two approaches. First, we include additional variables in the model and recover potential GDP using a minimal identification requirement with one set of constraints on the long-run effects of the shocks. The main idea is that since the potential growth builds upon supply shocks only, one set of constraints, stating that only supply shocks have permanent effects on the level of GDP in the long-run, is enough to recover the trend. We formally show that further restrictions are not needed as the other structural shocks are not interpreted and do not need to be separately identified. Furthermore, we show that the limited set of constraints used is the only one relevant for the decomposition of the output gap into contributions of the observables (see Andrle (2013) for a general explanation of the decomposition technique). With a different objective, similar semi-structural VARs have been used by Bernanke and Mihov (1998), King et al. (1991) and Gali (1999).

Second, in the spirit of Borio et al. (2017), we directly include pre-transformed financial cycle indicators in the model instead of estimating or imposing theory-based cointegrating relations between GDP and the financial variables. In other words, we condition the output gap estimates only upon the short- and medium-run correlations between GDP and the indicator variables, while the fundamental, long-run relationships between GDP and financial variables are not explicitly modelled. As with the model of Borio et al. (2017), this "shortcut" allows us to keep the dimension of the model relatively small. On the other hand, in contrast to Borio et al. (2017), financial cycle indicators are treated as stochastic processes, i.e. shocks can affect the indicators without necessarily influencing the GDP.

Since the semi-structural VAR involves entirely different mechanisms to decompose the trend from the cycle than the two other UCMs, it can provide particularly informative additional insights compared to the other models. All the more so as the ability of the usual univariate or multivariate statistical filters to accurately differentiate between supply and demand shocks has recently been challenged. Coibion et al. (2018) argue that models relying on smoothing techniques gradually but persistently respond to all kinds of shocks to the GDP. The authors show that the Blanchard and Quah (1989) approach, which explicitly distinguishes between temporary and permanent shocks, does not suffer from this shortcoming, and can generate real-time estimates of potential output that are consistent with theoretical predictions much more successfully.

As a first step, we estimate the three "baseline" models for eight advanced economies – Canada (CA), France (FR), Germany (DE), Italy (IT), Japan (JP), Spain (ES), the UK, and the US – in which the cyclical pattern of the demand is captured by the unemployment rate, the capacity utilisation rate, and (in the semi-structural UCM) CPI inflation. As a second step, we test the implications of adding two additional variables that proved to be the most successful in capturing financial cycles: credit and house prices (see e.g. the literature review by Borio (2014)). The models are evaluated along several dimensions, such as the contribution of financial shocks to the estimated output gaps, the sensitivity of the results to pre-treatment methods, or the real time performance of the models. Finally, the overall accuracy of the estimated output gaps are assessed using receiver operating characteristic analysis based on their capabilities in predicting recessions.

The overall picture underlines the importance of taking financial variables into account when assessing the cyclical position of the economy. Independently of the model considered, the model augmented with financial variables proves to be consistently more effective in identifying unsustainable economic growth paths and in predicting recessions both in-sample and out-of-sample. In ES, the UK and (to a much lesser extent) the US, the credit and house prices boom in the run-up to the Great Recession is clearly identified, as well as a previous boom-bust in the UK during the 1980s-90s. In JP both credit and house prices booms at the turn of the 1990s led to the well-studied and prolonged crisis (the "Lost Decade"). In FR the models signal a house prices boom during the 2000s, but there is no sign of credit boom during the estimation period. Finally, the financial cycles appear relatively muted in CA, DE and IT. At the same time, our results suggest that financial information generally worsens the real-time stability of the models in both the short-run and the longer-run, sometimes even for countries without clear financial cycles identified ex-post (most notably for DE). In other words, the models with financial variables are generally more successful in signalling booms in real-time, but future revisions of the point estimates are larger.

The models under scrutiny have different properties along several important dimensions, which affect their practical efficiency in handling new information originating from the financial variables and thus their resulting finance-neutral output gap estimates. The approach proposed by Borio et al. (2017) is particularly appealing for its simplicity and real-time stability. Since the additional variables directly enter the output gap equation as deterministic covariates, there are only a few parameters to estimate and shocks to decompose and, as a consequence, the resulting output gap estimates are relatively less revised when new data become available. However, the extended HP filter generally lags behind the two semi-structural models in terms of forecasting performance, in particular when the models are augmented with financial variables. Moreover, the reduced form approach is sensitive to the (necessary) pre-transformation applied to the cycle indicators to make them stationary with zero mean. Without external validation of the resulting output gap estimates, it is often hard to choose between the different possible trend removal methods. Simply put, the extended HP filter can be viewed as the two-step version of the more complicated, semi-structural UCM: first, possible structural shocks to the indicator variables need to be removed; second, the correlation between the pre-filtered cycle indicator variables and the GDP is estimated, which helps to identify the output gap.

The semi-structural UCM provides a simultaneous estimation of these two steps. The difficulty of selecting the right pre-treatment method is therefore transformed into a difficulty of choosing the right structure and defining suitable Bayesian prior distributions on the parameters. By imposing a very light structure on the model and using standard prior beliefs about the parameters, the semi-structural UCM produces, in general, more accurate finance-neutral output gaps with better early warning capabilities than the extended HP filter. However, the results for JP and – to a lesser extent

– the US are good examples showing that solely relying on an arguably flexible and generally more capable approach may also lead to misleading conclusions. The semi-structural UCM for these two countries reveals significantly lower impact of financial cycles on the output gap estimates compared to the other models. In particular, both the semi-structural VAR and the reduced form model suggest that the Japanese finance-neutral output gap remains positive until 1998 following the asset price bubble's collapse in the early 1990s, whereas the output gap estimated using the semi-structural UCM already turns to negative – even though only temporarily – in 1993. The output gap remaining positive for a prolonged period of time after the bubble's collapse is arguably more in line with the Japanese "Lost Decade" paradigm. Similarly, the build-up of the financial bubble in the US was less forceful before the recent financial crisis according to the semi-structural UCM than the other models imply.

By taking advantage of its distinct trend-cycle decomposition technique, the semi-structural VAR provides valuable external validation of the results from the other models. This cross-check is particularly useful when the results from the other UCMs differ (such as for JP and the US) or the results are particularly sensitive to the priors imposed on the parameters of the model. Moreover, the semi-structural VAR is generally more successful in accurately capturing macroeconomic and financial imbalances than the UCMs. The model features the best in-sample forecasting performance of recession probabilities among all three models, and it has similar real-time prediction capabilities to the semi-structural UCM and clearly superior to the extended HP filter. At the same time, this model also tends to react relatively more erratically to new observations included in the sample, especially when financial cycles are taken into account. The semi-structural VAR is also sensitive to the trend removal technique applied to the cycle indicators. However, contrary to the reduced form extended HP filter, the implied change in the output gap estimate following a change in the pre-filtering technique is less predictable.

In what follows, the two UCMs are briefly described (Section 4.2), and then the semi-structural VAR approach is presented in more detail (Section 4.3). Section 4.4 presents the data used for the estimations, the main estimation results, the decomposition of the output gap into the contribution of observables, the real-time performance and the forecasting accuracy analysis of the models. Finally, Section 4.5 summarises the main findings and puts them into a broader perspective.

4.2 The unobserved components models

4.2.1 Extended Hodrick-Prescott filter

Our simplest model has been advanced by Borio et al. (2017). A simple UCM, a univariate smooth trend model, which is the state-space representation of the famous HP filter, is augmented with additional variables (z_t):

$$\begin{cases} y_t = y_t^\star + \gamma' z_t + \varepsilon_t & \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \\ y_t^\star = y_{t-1}^\star + g_{t-1}^y \\ g_t^y = g_{t-1}^y + \xi_t & \xi_t \sim N\left(0, \sigma_\xi^2\right) \end{cases} \tag{4.1}$$

where y_t is the log of GDP, $y_t^\star$ is the log of the unobserved GDP trend and g_t^y is the stochastic growth rate of the trend. ε_t and ξ_t are uncorrelated normally distributed random terms with variance σ_ε^2 and σ_ξ^2, respectively. The output gap is defined as the deviation of GDP from its sustainable path: $\hat{y}_t = y_t - y_t^\star$. The parameters of the model are estimated using maximum likelihood techniques and the gap is recovered using the Kalman filter.

A number of issues need to be addressed. To begin with, the stationarity of the business and financial cycle indicators (vector z_t) is a crucial yet in practice a rarely satisfied assumption. The coefficient of a non-stationary variable z_t in small samples can receive a positive weight that will transmit a trend (or a non-zero mean) to the output gap. To overcome this problem, we follow the usual approach and pre-transform the series before plugging them into the model. A number of academic studies researching on financial imbalances de-trend the series using a HP filter with a smoothing parameter $\lambda = 400,000$ (see e.g. Borio and Lowe (2002); Edge and Meisenzahl (2011); Detken et al. (2013); Borio et al. (2014a); Anundsen et al. (2016); Bauer and Granziera (2017); and the Basel III recommendation Committee (2010)). Consequently, only very low frequency cycles with average length above approximately 40 years are removed, well above the usual financial cycle frequencies.[3] We follow the same procedure and de-trend the unemployment rate (u_t), the credit (cr_t) and the house prices (hp_t) using a HP filter with the same value for λ. There is one exception: for DE, a HP filter with a lower smoothing parameter ($\lambda = 1600$) is used to better account for the downward trend of the unemployment rate following the introduction of the Hartz reforms. As for the capacity utilisation (c_t), we simply remove the mean. The sensitivity of the results to various alternative low-frequency trend removal techniques will be discussed further in Sections 4.4.4 and 4.5.[4]

The second issue is the usual "pile-up" problem when maximum likelihood technique is employed: when the variation of trend growth rate σ_ξ^2 is small, the (maximum likelihood) estimate for σ_ξ^2 tends to be biased toward zero and as a result, the filter smooths potential output more than is necessary. The business cycle would then seem longer (Shephard (1993); Stock (1994); Stock and Watson (1998)). To avoid this, we follow the suggestion of Borio et al. (2017) and constrain the scaling factor $\lambda = \sigma_\varepsilon^2 / \sigma_\xi^2$ so that the ratio of the variance of the output gap to the variance of the second differences of the trend –

[3]The transfer function of the HP filter is given by (King and Rebelo (1993)): $h(\omega;\lambda) = 4\lambda(1-\cos\omega)^2/(1+4\lambda(1-\cos\omega)^2)$. If $\lambda = 400,000$, $h(\omega;\lambda) > 0.5$ for $\omega > 0.04$, which corresponds to about 39 years on quarterly data.

[4]Note that we did not seek to fully optimise the pre-treatment of the variables to achieve the most "credible" output gap estimates. Instead, we used rather standard pre-treatment techniques to illustrate the potential benefits and pitfalls of the different approaches.

i.e. $\text{var}(\hat{y}_t)/\sigma_\xi^2$ – is equal to 1600, the standard value for quarterly data. This is achieved by recursively re-estimating the model until convergence.

Finally, the model suffers from two conceptual weaknesses. First, the extended HP filter might be too restrictive. The smooth trend model in eq. 4.1 with the above-mentioned constraint on the scaling factor assumes that the trend follows an integrated random walk. As such, the model does not allow for a level shock on potential output, since it allocates all permanent shocks to the potential output as shocks to the trend growth. Moreover, the model does not assume autocorrelation of the output gap. However, in practice, unit root tests on real GDP rather indicate integrated processes of order 1 (i.e. an I(1) process), rather than I(2) processes as assumed by the smooth trend model; and the output gap estimates produced by the filter are highly autocorrelated, which contradicts the formulation of the model (Grant and Chan (2017)).

Second and most importantly from the point of view of our exercise, the extended HP filter method may not be compatible with filtering out financial imbalances from output to obtain sustainable output because of the frequency discrepancy between financial and regular economic cycles. Using the usual value for the scaling factor, the extended HP filter is designed to filter out noise at the traditional business cycle frequency. Additional variables in z_t will be significant determinant of the output gap only if they can explain cyclical fluctuations at the traditional business cycle frequency (as, for example, the unemployment and the rate of capacity utilisation). However, financial variables have longer cycles and their correlation with GDP is expected at lower frequencies. A common business cycle covers 8 years, whereas on average the financial cycle is considerably longer, 15-20 years (see Claessens et al. (2012); Borio (2014); Rünstler and Vlekke (2018)).

To address these conceptual issues, we also estimate a more flexible version of the model which (*i*) allows for the serial correlation of the output gap, (*ii*) introduces shocks to the level of the trend and (*iii*) estimates the scaling factor rather than calibrates it. This model is similar in structure to the process describing the potential output and the output gap in the semi-structural UCM (model B in Section 4.2.2). It is also more in line with the assumptions used in the semi-structural VAR (model C in Section 4.3), in particular regarding the integration order of the GDP series. The model is described in details in Appendix 4.B ("dynamic multivariate filter", model A′). The results from these two versions of the reduced form model are very similar: the correlation between the output gap estimates using these two methods, with or without financial variables, is very high, that is, more than 0.89 for all countries. In other words, the additional flexibility of the dynamic multivariate filter compared to extended HP has very limited practical relevance. Therefore, the results of the more flexible reduced form model are reported only in Appendix 4.B.

4.2.2 Semi-structural unobserved components model

Our baseline model is similar to Melolinna and Tóth (2019), Anderton et al. (2014) (ECB), Benes et al. (2010) (IMF), and IMF-QPM model (Ermolaev et al. (2008), following Kuttner (1994)). It is a backward looking UCM with a Phillips curve, an Okun's law, and a stochastic process relating output gap to capacity utilisation. The baseline model includes four observable variables: log GDP (y_t), inflation (π_t) defined as the yearly log difference of the consumer price index (CPI), unemployment (u_t) and capacity utilisation (c_t).

Consequently, the model has four measurement equations with x_t being the observable variable, $x_t^\star$ the unobservable trend and $\hat{x}$ the gap:

$$x_t = x_t^\star + \hat{x}_t, x = \{y, \pi, u, c\} \tag{4.2}$$

The decomposition of output into its potential and the output gap is described by the following three equations. The output gap is assumed to follow an AR(1) process, therefore featuring a serial correlation and experiencing transitory shocks. Potential output is defined as an I(1) process. Its stationary growth rate is an AR(1) revolving around its sample average. ε^{y^*} and ε^{gy} are permanent and temporary shocks to the potential output, respectively.

$$\begin{cases} \hat{y}_t = \alpha_1 \hat{y}_{t-1} + \varepsilon_t^{\hat{y}} \\ y_t^\star = y_{t-1}^\star + g_t^y + \varepsilon_t^{y^\star} \\ g_t^y = \alpha_2 \bar{g}^y + (1-\alpha_2)\, g_{t-1}^y + \varepsilon_t^{gy} \end{cases} \tag{4.3}$$

An Okun's law links unemployment and output gap. The unemployment gap includes a serial correlation and is supposed to be negatively correlated to the output gap. Potential unemployment is an I(1) process. Its first-difference follows an AR(1) process with zero mean.

$$\begin{cases} \hat{u}_t = \gamma_1 \hat{u}_{t-1} - \gamma_2 \hat{y}_t + \varepsilon_t^{\hat{u}} \\ u_t^\star = u_{t-1}^\star + g_t^u + \varepsilon_t^{u^\star} \\ g_t^u = (1-\gamma_3)\, g_{t-1}^u + \varepsilon_t^{gu} \end{cases} \tag{4.4}$$

An equation describing the relationship between capacity utilisation and output gap, in the spirit of the Okun's law, provides further identifying information. The capacity utilisation gap is assumed to be persistent and correlated with the output gap. Potential capacity utilisation is modelled as a driftless random walk.

$$\begin{cases} \hat{c}_t = \kappa_1 \hat{c}_{t-1} + \kappa_2 \hat{y}_t + \varepsilon_t^{\hat{c}} \\ c_t^\star = c_{t-1}^\star + \varepsilon_t^{c^\star} \end{cases} \tag{4.5}$$

A Phillips curve equation links output gap and inflation. The inflation gap is persistent (it is modelled as an AR(1) process) and is correlated with the output gap. Potential inflation is assumed to be a random walk without drift.

$$\begin{cases} \hat{\pi}_t = \beta_1 \hat{\pi}_{t-1} + \beta_2 \hat{y}_{t-1} + \varepsilon_t^{\hat{\pi}} \\ \pi_t^\star = \pi_{t-1}^\star + \varepsilon_t^{\pi} \end{cases} \tag{4.6}$$

The inclusion of the Phillips curve in the semi-structural model arguably provides additional information that is not present in the other models (A and C) tested. We add the Phillips curve to this model to be consistent with other similar UCMs used in the literature. However, we equally tested the model without the Phillips curve. Results from this model variant remain very close to our baseline findings. Similarly, we also tested a variant of the other models with inflation as an additional control. Our results confirm the findings of Borio et al. (2017) demonstrating that inflation "does very little to help condition output gap estimates". These additional results are available upon request.

The model with financial indicators corresponds to the baseline model augmented with financial variables $f_t = \{cr, hp\}$. They were added separately, before they were added together. First, a measurement equation decomposes the log of credit (cr) and the log of house prices (hp) into a trend and a cycle components:

$$f_t = f_t^\star + \hat{f}_t, f \in \{cr, hp\} \tag{4.7}$$

Second, the trend-cycle decomposition of f_t is similar to that for output. The financial variable gap is linked to the output gap and is persistent. The trend of the financial variable is defined as I(1). Its stationary growth rate is an AR(1) process revolving around the sample average of f.

$$\begin{cases} \hat{f}_t = \phi_1 \hat{f}_{t-1} + \phi_2 \hat{y}_t + \varepsilon_t^{\hat{f}} \\ f_t^\star = f_{t-1}^\star + g_t^f + \varepsilon_t^{f^\star} \\ g_t^f = \phi_3 \bar{g}^f + (1 - \phi_3) g_{t-1}^f + \varepsilon_t^{gf} \end{cases} \tag{4.8}$$

The model is estimated using a diffuse Kalman filter (Rosenberg (1973); De Jong (1991)) and Bayesian techniques.[5] Shocks are assumed to follow an inverse gamma function with infinite variance. The AR(1) parameters follow a beta distribution with prior mode 0.6 and prior variance of 0.25. The rest of the priors follow gamma, beta or normal distributions. Details are presented in Appendix 4.C.

[5]We use the IRIS Toolbox, see: J. Benes, M. K. Johnston, and S. Plotnikov, IRIS Toolbox Release 20151016 (Macroeconomic modelling toolbox). The software is available at http://www.iris-toolbox.com

4.3 Semi-structural VAR with long-run restrictions

Building on the seminal article by Blanchard and Quah (1989), we impose long-run restrictions to disentangle shocks to the trend influencing GDP in the long-term from shocks to the cycle with no influence on long-term GDP. We extend the original Blanchard-Quah model in two ways. First, instead of relying on two variables (GDP and an indicator of the business cycle such as the unemployment rate) and one long-run restriction (only the supply shock has a permanent effect on the level of GDP), we exploit a wider set of information by using several cycle indicators without imposing further restrictions. To avoid a fragile identification of all shocks through various long-run, short-run and/or sign restrictions, Blanchard and Quah (1989)'s original approach is used and only the same long-run constraints are imposed: only one of the structural shocks (the supply shock) impacts the GDP in the long run. This minimal identification requirement is sufficient to recover the supply shock, the potential growth, and thus the output gap. The other structural shocks (for simplicity, we call them demand shocks) are not interpreted.

Since the entire structure of the shocks is not identified, but only one of these shocks, the identification technique relates to semi-structural VARs (see Kilian and Lütkepohl (2017), Chapter 10). Similar semi-structural VARs in which only part of the structural shocks are correctly identified were used by, for example, Bernanke and Mihov (1998) to derive a new measure of monetary policy innovations based on various reserve market indicators; or King et al. (1991) to identify common permanent productivity shocks in output, consumption, and investment based on the "balanced growth" assumption. Our approach is the closest to the higher dimension model of Gali (1999), who identify technology shocks driving the productivity growth by using information on hours worked, money growth, inflation, and interest rates. Nevertheless, use of semi-structural VAR in the empirical literature remains scarce and, to our knowledge, it has never been used to recover the output gap.

Second, in the spirit of Borio et al. (2017), we directly include pre-treated cycle indicators in the model. This "shortcut" is particularly convenient when non-stationary financial variables are included in the model, since it allows us to avoid the need to estimate or impose cointegrating relationships between GDP and the financial variables. In practice, we use the same pre-transformed series as in our "extended HP filter" model (Section 4.2.1).[6] However, in contrast to Borio et al. (2017), both the business and the financial cycle indicators enter the model as stochastic processes. Consequently, various shocks can affect the cycle indicators in the short and medium run without necessarily influencing the GDP. In other words, as opposed to the extended HP filter, shocks to the indicator variables are not necessarily directly transmitted to the output gap.

[6] Blanchard and Quah (1989) also stress the necessity of some form of pre-treatment applied to the indicator series. The authors discuss how to deal with the time trend in unemployment. They compare the results obtained with raw or detrended unemployment series, and show that the results obtained using either of the two series are qualitatively similar.

More formally and starting from the moving average (Wold) representation for the vector composed of growth Δy_t and (stationary and demeaned) auxiliary variables z_t (u_t, c_t, and in the augmented model, the two financial indicators: cr_t and hp_t), the structural shocks $[\omega_t^s, \omega_t^d]$ need to be recovered from the innovations ε_t:

$$\begin{bmatrix} \Delta y_t \\ z_t \end{bmatrix} = C(L)\,\varepsilon_t = A(L)\,\omega_t = A(L)\begin{bmatrix} \omega_t^s \\ \omega_t^d \end{bmatrix} \tag{4.9}$$

Blanchard and Quah (1989) only use information from one macroeconomic variable beyond GDP: in their model z_t is a variable (unemployment) instead of a vector. Therefore, they separate two shocks into a demand disturbance (ω_t^d) and a supply disturbance (ω_t^s). To achieve this, one constraint is enough: the demand disturbance has no effect on GDP in the long run.

In contrast, we separate one supply shock from many other demand shocks by using several indicator variables (z_t and ω_t^d are column vectors with n elements). Noting $A(L) = \sum_{i=0}^{\infty} A_i L^i$ and $C(L) = \sum_{i=0}^{\infty} C_i L^i$, and given that $C_0 = I$, then $A_0\omega_t = \varepsilon_t$ and $A_i = C_i A_0$. Since $E(\omega_t) = 0$, $E(\omega_t\omega_t') = I_{n+1}$ and $\varepsilon_t\varepsilon_t' = A_0\omega_t\omega_t'A_0'$, the variance-covariance matrix $V(\varepsilon)$ is given by:

$$V(\varepsilon) = A_0 A_0' \tag{4.10}$$

$A(1)\omega_t = C(1)A_0\omega_t$ is the long-term accumulated response of the GDP growth and the auxiliary variables to the structural shocks ω_t. It can therefore be imposed that only the supply shock (that we choose as the first component of ω_t) has an impact on the level of GDP in the long run:

$$A(1) = C(1)A_0 = \left[\begin{array}{c|c} a_{11} & 0_{1\times n} \\ \hline A_{n\times 1} & Z_{n\times n} \end{array}\right] \tag{4.11}$$

The n zeroes from the first line in eq. 4.11 impose that only the supply shock ω_t^s is allowed to impact GDP in the long-run. These are the constraints that matter to separate the supply shock from the other shocks. We further impose $n \times (n-1)/2$ arbitrary constraints on the matrix Z to be able to identify A_0 and therefore all of the structural shocks (e.g. by imposing Z to be lower triangular),[7] but crucially the potential growth (i.e. growth stemming from the supply shock only) and therefore the output gap do not depend on these constraints. To put it differently, the constraints on Z split the innovations between the various demand shocks, but the output gap is the same whatever the construction of the demand shocks. For the same reason, the contribution of each variable to potential growth does not depend on these additional restrictions set on Z: this contribution reflects the part

[7]The identification of A_0 requires $(n+1)^2$ constraints. Equation 4.10 yields $(n+1) \times (n+2)/2$ restrictions, and the first line of $A(1)$ in eq. 4.11 imposes an additional n constraints. The remaining $n \times (n-1)/2$ exclusion restrictions are placed on the matrix Z.

of the variable having an impact on long-term GDP. How the rest of this variable – the part with no impact on long-term GDP – is split between the different demand shocks is therefore irrelevant, as shown below.

Let's denote $\begin{bmatrix} \omega_t^s \\ 0_{n\times 1} \end{bmatrix} = \left[\begin{array}{c|c} 1 & 0_{1\times n} \\ \hline 0_{n\times 1} & 0_{n\times n} \end{array}\right] \omega_t = B\omega_t$. $y_t^\star$ is estimated by setting all structural shocks ω_t^d to 0:

$$\begin{aligned}
\begin{bmatrix} \Delta y_t^\star \\ z_t^\star \end{bmatrix} &= A(L)B\omega_t = C(L)A_0B\omega_t = C(L)A_0BA_0^{-1}\varepsilon_t = C(L)C(1)^{-1}A(1)BA(1)^{-1}C(1)\varepsilon_t \\
&= C(L)C(1)^{-1} \left[\begin{array}{c|c} a_{11} & 0_{1\times n} \\ \hline A_{n\times 1} & Z_{n\times n} \end{array}\right] \left[\begin{array}{c|c} 1 & 0_{1\times n} \\ \hline 0_{n\times 1} & 0_{n\times n} \end{array}\right] \left[\begin{array}{c|c} \frac{1}{a_{11}} & 0_{1\times n} \\ \hline -\frac{1}{a_{11}} Z_{n\times n}^{-1} A_{n\times 1} & Z_{n\times n}^{-1} \end{array}\right] C(1)\varepsilon_t \\
&= C(L)C(1)^{-1} \left[\begin{array}{c|c} 1 & 0_{1\times n} \\ \hline \frac{1}{a_{11}} A_{n\times 1} & 0_{n\times n} \end{array}\right] C(1)\varepsilon_t
\end{aligned} \tag{4.12}$$

$C(L)$ and $C(1)$ (inverse of the VAR) depend only on reduced-form parameters of the VAR. Futhermore, the parameters a_{11} and $A_{n\times 1}$ in $A(1)$ are identified by the equations 4.10 and 4.11 without requiring the knowledge of Z:

$$\begin{aligned}
A(1)A(1)' &= \left[\begin{array}{c|c} a_{11}^2 & a_{11}A'_{1\times n} \\ \hline a_{11}A_{n\times 1} & A_{n\times 1}A'_{1\times n} + Z_{n\times n}Z'_{n\times n} \end{array}\right] \\
&= C(1)A_0A_0'C(1)' = C(1)V(\varepsilon)C(1)'
\end{aligned} \tag{4.13}$$

Given that none of the matrices involved in the last line of eq. 4.12 depend on Z, the estimation of potential growth $\Delta y_t^\star$ – and that of $z_t^\star$ – is independent of the exact specification of Z.

We then compute the output gap $\hat{y}_t$ as $y_t - y_t^\star = \sum_{t_0=1}^t \Delta y_t - \sum_{t_0=1}^t \Delta y_t^\star - \overline{\sum_{t_0=1}^t \Delta y_t - \sum_{t_0=1}^t \Delta y_t^\star}$, with $\overline{\sum_{t_0=1}^t \Delta y_t - \sum_{t_0=1}^t \Delta y_t^\star}$ being the mean of the cumulated difference between growth and potential growth. This latter term ensures that the output gap has mean zero. The output gap $\hat{y}_t$ does not depend on the matrix Z because neither $\sum_{t_0=1}^t \Delta y_t$ nor $\sum_{t_0=1}^t \Delta y_t^\star$ depends on it.

Finally, note that the contribution of one variable to the output gap is therefore simply the cumulative difference between its contribution to growth and its contribution to potential growth. Replacing ε_t with $C(L)^{-1}[\Delta y_t, z_t]'$ in eq. 4.12, it can be seen that the decomposition of the output gap into the observables does not depend on the matrix Z either.

This semi-structural VAR has some appealing advantages: it imposes a very light structure on the data, thus “lets the data speak”. In particular, since we do not need to identify all structural shocks, the identifying assumption requires only minimal restrictions. Although the series must be stationary and must therefore be transformed before the estimation (we perform the same pre-transformation as in the reduced-form models), the auxiliary variables are still treated as stochastic processes. Finally the estimation is straightforward: we use the Full Information Maximum Likelihood (FIML) procedure and the scoring method described in Amisano and Giannini (1997).

4.4 Results

4.4.1 Data source

We use quarterly and seasonally adjusted data for eight advanced economies: CA, FR, DE, IT, JP, ES, the UK and the US. We took the longest publicly available series for each country. Start dates range from 1968q4 to 1985q1 depending on the country, while the end date is always 2016q4.[8] Gross domestic product, consumer price index and the rate of capacity utilisation are obtained from the OECD Main Indicators database. The unemployment rate comes from the OECD Economic Outlook database. Total credit to the non-financial sector comes from the BIS total credit statistics database. Real house prices are calculated using house prices from national sources, BIS Residential Property Price database and the Consumer Price Index from the OECD.[9] The credit and house prices series were seasonally adjusted. Further details on the data can be found in the Appendix.

4.4.2 Main results

Figure 4.1 shows, for each country and each model, the estimated output gaps without financial information (baseline, red line), and augmented with credit (blue line), house prices (green line) or both (black line). The differences between the augmented and the baseline models are the impacts of introducing financial cycle indicators into the models on the output gap estimates. As shown in more detail in the next section, these differences – with some exceptions outlined in the paper – are highly correlated with the contributions of the financial variables to the output gap. Therefore, the differences in the output gap estimates with and without financial variables can safely be interpreted as signs of financial imbalances not reflected in the baseline output gaps. We start by discussing and comparing

[8]The estimation period begins in 1968q4 for IT, 1970q1 for CA, DE and the US, 1976q1 for FR, 1976q1 for ES, 1978q1 for JP and 1985q1 for the UK.

[9]To get longer time series, we use other data sources for the rate of capacity utilisation for CA and JP. See Appendix 4.A for details.

the results obtained using the two UCMs, then we turn our attention to the semi-structural VAR and show how it shapes our understanding of financial cycles and their interactions with GDP.

The graphs clearly show that the baseline output gap estimates from the two UCMs are close to each other for all countries.[10] The models generally produce economically plausible output gap estimates. The economic downturn starting with the 2007-08 financial crisis – and further aggravated by the subsequent debt crisis for IT and ES – appears as the worst crisis in our sample for most countries. The only exception is CA, for which it turned out to be relatively benign, especially so in comparison with its severe crisis in the 1980s.

Adding financial information generally drives the two UCMs' output gap estimates apart. Most notably, the UCMs seem to use the additional information provided by the financial variables differently in the case of JP: the extended HP filter reveals a significantly higher and more prolonged impact of the Japanese financial bubble of the 1990s than the semi-structural UCM. The former model suggests that the finance-neutral output gap remains positive until 1998, whereas the latter model already indicates – even though only temporarily – a negative output gap in 1993.

The results from the UCMs are also divergent for the UK: while the semi-structural UCM clearly shows the impact of both credit and housing bubbles in the late 1980s and in the 2000s, the extended HP filter only signals the effects of house prices cycles. Similarly, the extended HP filter points towards a somewhat more important role of housing bubbles in FR and credit bubbles in DE during the years preceding the crisis.

Turning to the semi-structural VAR, results show no sign of systematic differences between this model and the UCMs that could be summarized in a single, clear message. Instead, the example of the 8 countries illustrate distinct cases, highlighting specific features of the semi-structural VAR and diverse aspects of its contribution to the overall picture.

Although the baseline semi-structural VAR draws a similar picture than the UCMs for most countries, there are also some marked differences. In particular, the model indicates a higher pre-crisis output gap compared to the other baseline models for FR and ES. Conversely, the semi-structural VAR without financial variables seems to be unable to pick up the high volatility in output around the outbreak of the crisis in the UK and it generates relatively less informative high frequency output gap fluctuations for DE.

The model with financial variables reinforces the findings from the extended HP filter and somewhat contradicts the findings from the semi-structural UCM in the case of JP and (to a lesser extent) the US. More precisely, both the semi-structural VAR and the reduced form model suggest that the Japanese finance-neutral output gap remains positive for a prolonged period of time after the asset

[10]The correlation between the output gaps from the extended HP filter and the semi-structural UCM exceeds 0.82 for all countries.

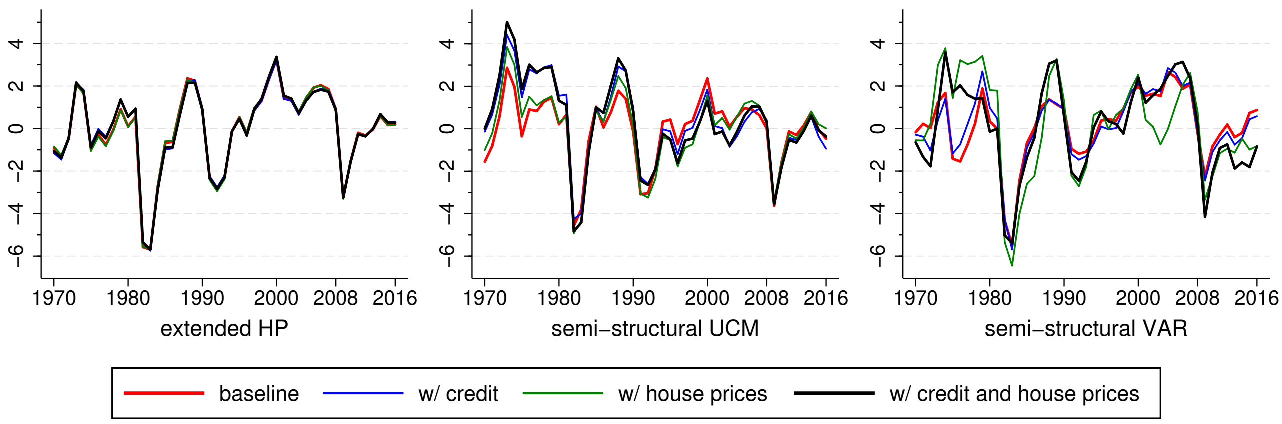
price bubble's collapse in the early 1990s, more in line with the Japanese "Lost Decade" paradigm. Similarly, the build-up of the financial imbalances in the US before 2007 seems to be more forceful than the semi-structural UCM implies.

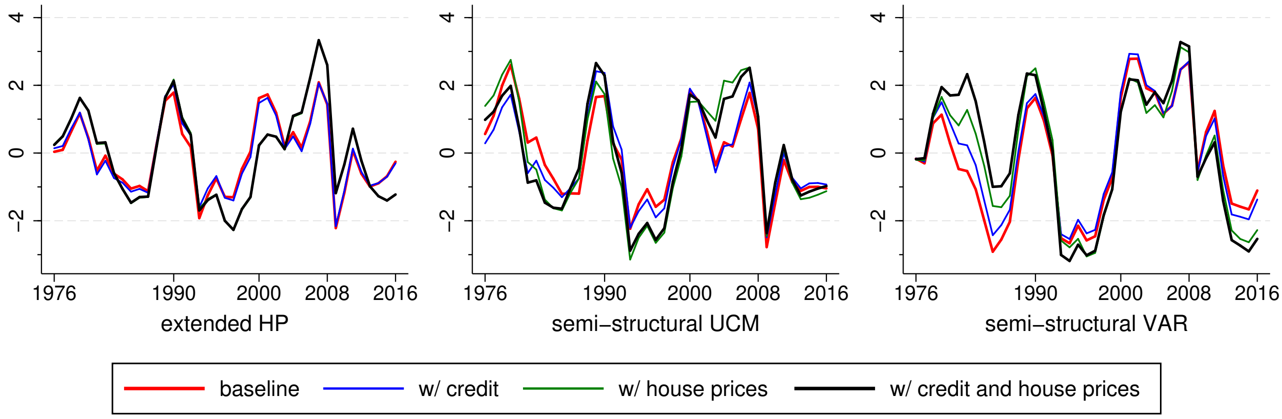
On the other hand, the semi-structural VAR gives credit to the semi-structural UCM in the case of the UK, suggesting that both credit and house prices cycles have been important sources of macroeconomic fluctuations in the country. Finally, the semi-structural VAR gives a similar general picture on CA and IT to the other models, showing no clear sign of impact of financial cycles on GDP. The only country for which the semi-structural VAR provides less credible or informative output gap estimates compared to the other models, either with or without financial variables, is DE.

An overall picture emerges from the results of all the models combined. Our findings are in line with previous work that documents the exceptional resilience of the Canadian economy in the midst of the global financial crisis (Bordo et al. (2015); Haltom (2013)). The financial cycles appear to be relatively muted in DE (as in Rünstler and Vlekke (2018)) and in IT as well. The models signal a house prices boom in FR during the 2000s. Both credit and house prices boom in JP at the turn of the 1990s, leading to a well-studied prolonged crisis (see e.g. Bayoumi (2001)). In the US, the credit and house prices boom in the run-up to the Great Recession is clearly visible. Corroborating the findings of Rünstler and Vlekke (2018), our results indicate that ES and the UK have suffered from the largest and the longest financial cycles.

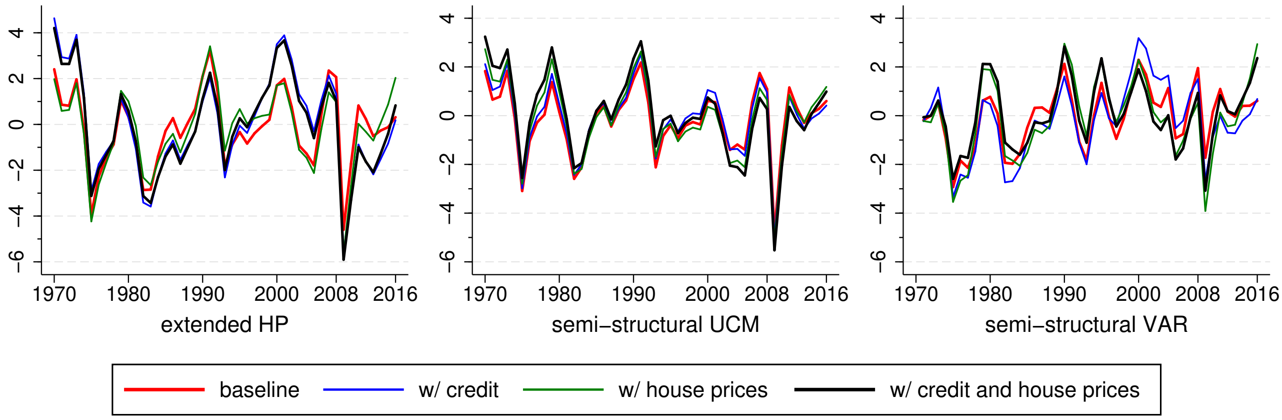
Figure 4.1: Output gap estimates

(a) Canada

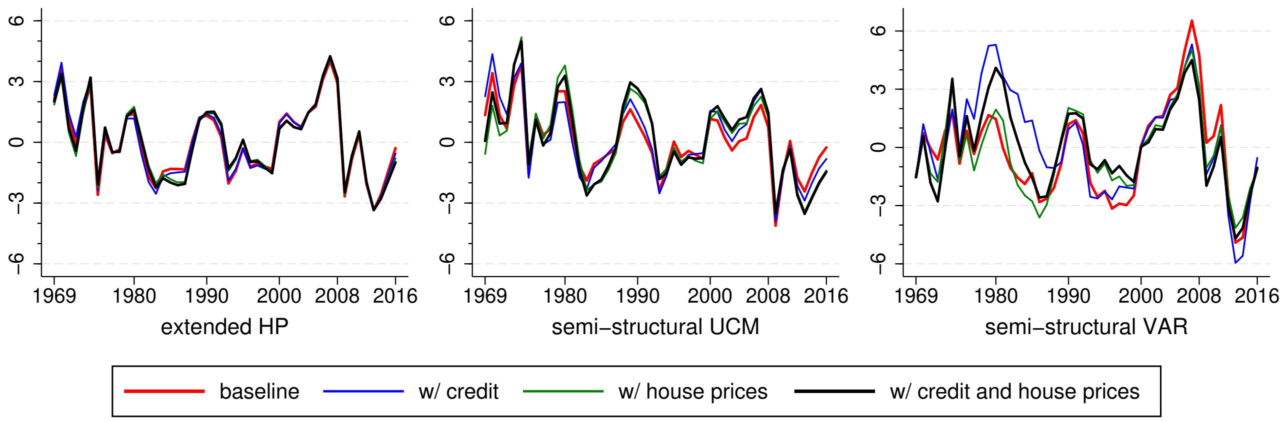




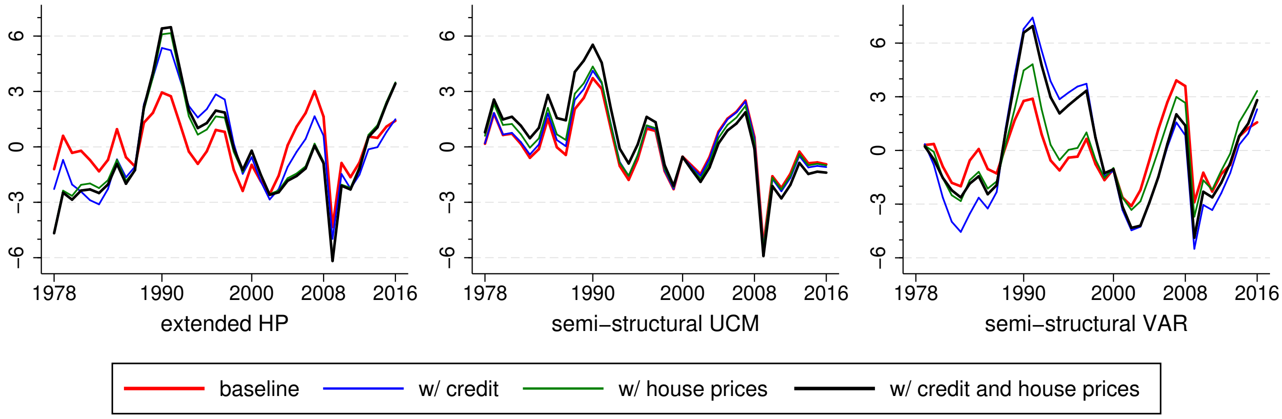
(b) France



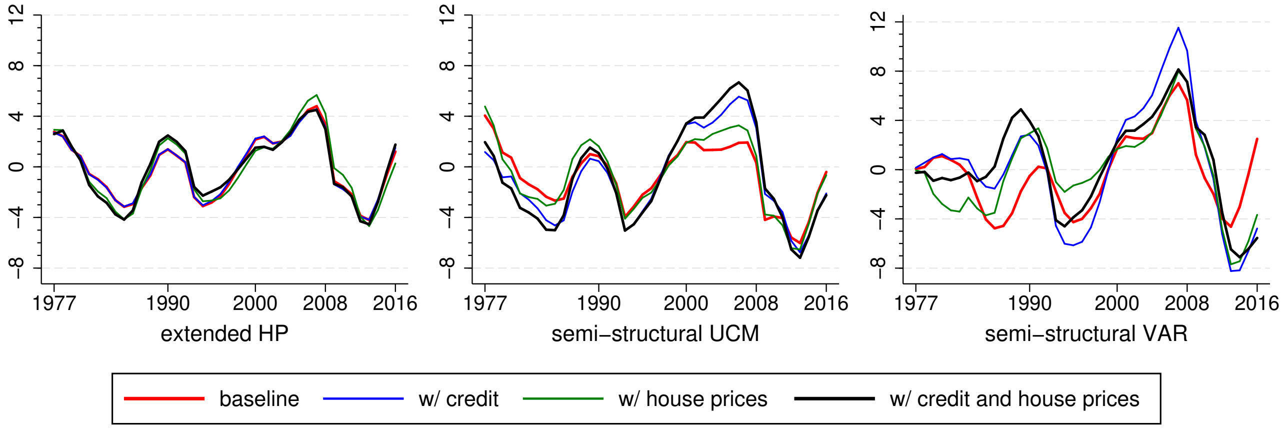
(c) Germany



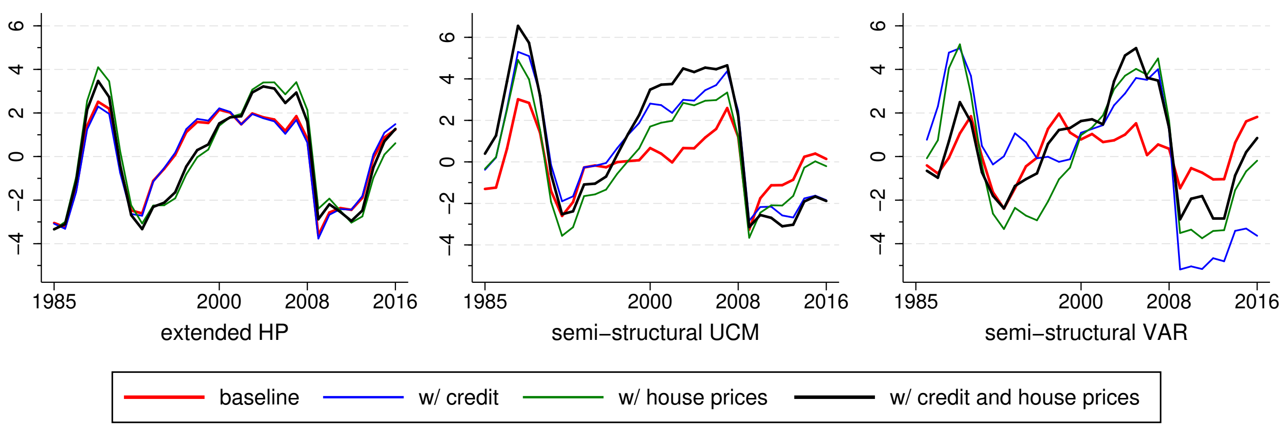
(d) Italy



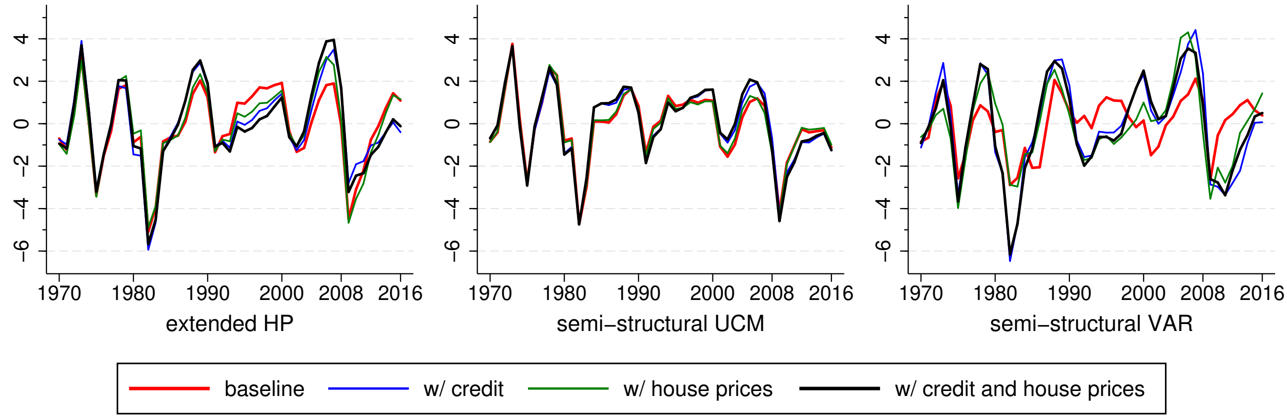
(e) Japan



(f) Spain



(g) UK



(h) US

Notes: Figures 4.1(a) to 4.1(h) display the output gap estimates for the extended HP filter (graph on the left), the semi-structural UCM (middle) and the semi-structural VAR (right) for each country. In each case the baseline output gap (red) is shown along with the output gaps estimated with the addition of credit (blue), house prices (green) or both credit and house prices (black).

4.4.3 Decomposition of the output gap into the contribution of observables

Changes in the output gap estimates triggered by the introduction of financial variables do not necessarily signal financial imbalances. For instance, the parameters of the model might change with a new variable in the model. As a result, the estimated output gap might change even if the direct influence of the variable on the GDP is limited or, inversely, the new output gap might be close to the baseline estimate even if the influence of the variable is important.

To assess the direct influence of the financial variables on the output gap, we decompose the gap estimates into the contribution of each variable. All of the models considered can be formulated as linear filters: the output gap can be written as a moving average of the observables. Therefore, the output gap can be decomposed exactly into the contribution of each observable. It is different from the more widely used shock decomposition: the contributions of the observables assess the influence of the variables themselves – not their shocks – on the estimated output gap. The technique is described in Andrle (2013). Section 4.3 gives additional details about the methodology applied to the semi-structural VAR.

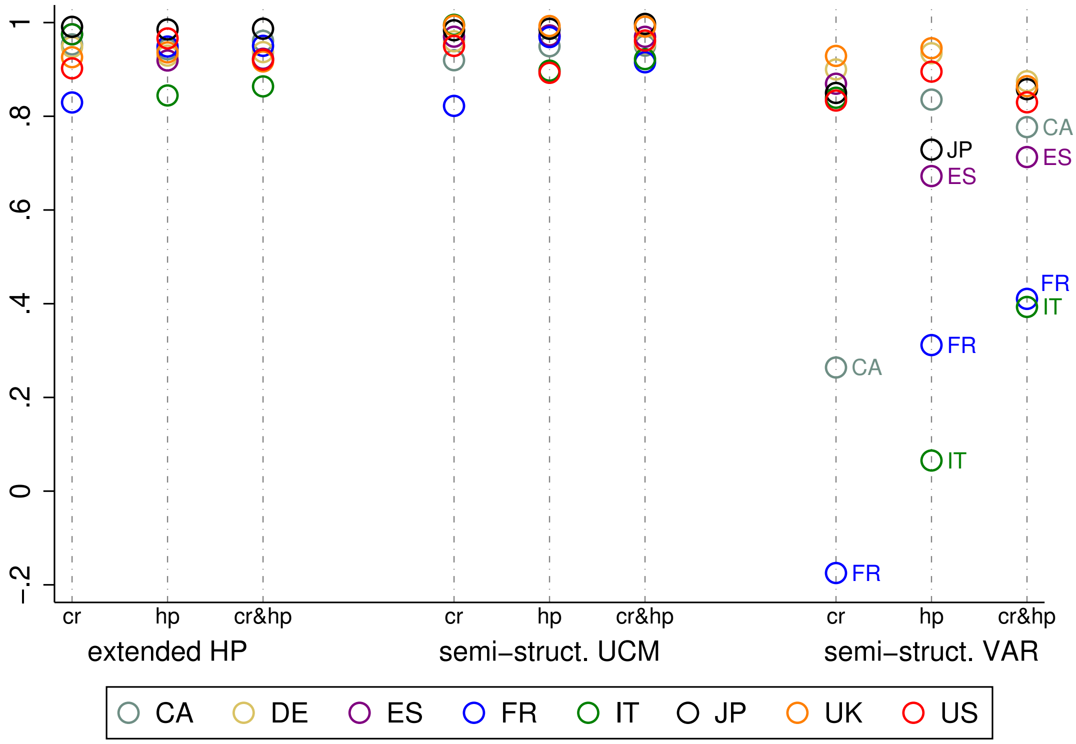
For each country and each model, we compare the contribution of credit (in the model with only credit), house prices (in the model with only house prices), or from both credit and house prices (in the model with both financial variables), with respectively the "credit deviation" (difference between the output gap from the model with only credit and the baseline output gap), the "house prices deviation", and the "credit and house prices deviation" (difference between the output gap from the model with both credit and house prices and the baseline output gap). Figure 4.2 describes the correlations for each country and model between these deviations from the baseline estimates and the contributions of the financial variables.

These correlations are generally very high, especially for the UCMs. The extremely low correlations that appear for the credit for CA and FR and for the house prices for IT in the case of the semi-structural VAR can be disregarded: the variables have virtually no influence on the output gap, so their contribution is close to 0, which makes the correlation rather meaningless (see Figures 4.1(a), (b) and (d)). On the other hand, the low correlation of the contribution of house prices and the house price deviation (as defined in the previous paragraph) for the semi-structural VAR for FR indicates that the baseline model is relatively less robust to the introduction of house prices and, therefore, the results should be interpreted with more caution. The same problem arises to a much lesser extent with the introduction of house prices into the semi-structural VAR model for ES and JP.

Overall, measuring the deviations from the baseline for the large majority of the countries and models is roughly equivalent to computing the contribution of the financial variable(s) to the output gap estimates. The differences between the augmented and the baseline estimates can therefore be interpreted as signs of financial imbalances not reflected in the baseline output gaps. For example, if the introduction of credit translates into a (much) higher output gap, the contribution from credit will be positive (and large). Both measures point to the same (large) credit boom.[11]

[11] The visual inspection of the graphs comparing the contribution of each financial variable and the deviation in the output gap due to the inclusion of the financial variable into the model confirms this finding. These graphs are not presented in this paper for brevity reasons, but are available from the authors upon request.

Figure 4.2: Correlation between deviations from the baseline and the contributions of financial variables



Notes: For each model and each country, we plot the correlation between (1) the contribution of credit ("cr", in the model with only credit), house prices ("hp", in the model with only house prices), or from both credit and house prices ("cr&hp", in the model with both financial variables) and (2) the respective "credit deviation" (difference between the output gap from the model with only credit and the baseline output gap), "house prices deviation", and "credit and house prices deviation". The correlation measures below 0.8 are labelled with the country code.

4.4.4 Sensitivity of the results to pre-treatment methods

The sensitivity of the results to the pre-transformation technique applied to the cycle indicators is tested by re-estimating both the extended HP filter and the semi-structural VAR with alternative pre-treatment methods. Instead of relying on HP-filtered variables, we de-trend each element of z_t (both the business cycle and financial cycle indicators) by regressing out the second or third order polynomial trend. As a third alternative, we feed the models with the cyclical part of the indicator variables obtained from the semi-structural UCM ($\hat{u}_t, \hat{c}_t, \hat{cr}_t, \hat{hp}_t$).

Table 4.1 presents the standardized mean deviations (SMDs) of the input variables and the resulting finance-neutral output gap estimates with respect to the reference model with HP-filtered indicators.

More precisely, we compute the average deviations as $\sum_{m=1}^{M}\sum_{t=t_0}^{T}\frac{1}{M(T-t_0+1)}\frac{|z_{it}^{m}-z_{it}^{\text{ref.}}|}{\sigma_i^{\text{ref.}}}$, with z_{it}^{m} are the indicator variables from the alternative pre-treatment method $m = \{1, ..., M\}$ of country i and $z_{it}^{\text{ref.}}$ are the reference HP-filtered indicators with standard deviations $\sigma_i^{\text{ref.}}$. SMDs of the corresponding output gaps are calculated in a similar way.

All countries and all alternative pre-filtering methods combined, the SMD of the estimated output gap with respect to the reference model (HP-filtered indicators) is 24 per cent of the overall standard deviation of the output gaps for the extended HP model and 50 per cent for the semi-structural VAR. These numbers correspond to an average deviation of 0.5 percentage points and 1.2 percentage points for the respective models. That is, the semi-structural VAR is more sensitive to the input indicators than the extended HP model.

The order of magnitude of the SMDs of the output gaps is in accordance with the SMDs of the indicator variables. On average, with an overall SMD of 0.58, the unemployment gap is the most affected by the choice of the pre-treatment method. This is especially true for DE, the country for which a lower smoothing parameter is used in the reference model in order to account better for the relatively rapid changes in the structural unemployment rate following the Hartz reforms (see Section 4.2.1). At the other extreme, the least impacted variable is the capacity utilisation.

The first two columns of Table 4.2 attempt to explain the differences in the results obtained using various pre-treatment methods by differences in cycle indicators using a simple linear model. We pool all countries and all pre-treatment methods (de-trended by second or third order polynomials; and the cycle indicators from the semi-structural UCM), we calculate the differences with respect to our reference method and regress the output gap differences on differences in the indicator measures using ordinary least squares. As shown in the first column, for the extended HP model, 66 per cent of the variation in the estimated output gaps can be explained by the variations in input indicators. These results indicate that the cycle indicators directly transmit to the estimated output gap to a large extent. This is not the case for the semi-structural VAR: while the model is also sensitive to the pre-treatment method applied, the change in the resulting output gap estimates is much less predictable. These results hold true when more lags are added to the model: even with 10 lags, the adjusted R2 is 0.68 in the case of the extended HP model and 0.02 in the case of the semi-structural VAR. When the parameters of the explanatory variables are allowed to vary by country, the adjusted R2 increases to 0.93 for the extended HP model and to 0.33 for the semi-structural VAR.[12]

[12]These additional results are not presented in the paper, but are available from the authors upon request.

Table 4.1: Standardised mean deviations w.r.t. the model with HP-filtered indicators

	$\hat{u}_t$	$\hat{c}_t$	$\hat{cr}_t$	$\hat{hp}_t$	$\hat{y}_t^{\text{extended HP}}$	$\hat{y}_t^{\text{semi-struct. VAR}}$
Canada	0.35	0.20	0.53	0.28	0.25	0.31
France	0.40	0.12	0.11	0.21	0.12	0.20
Germany	1.71	0.20	0.37	0.42	0.31	0.75
Italy	0.35	0.26	0.38	0.28	0.16	0.50
Japan	0.77	0.23	0.41	0.32	0.49	0.63
Spain	0.28	0.17	0.25	0.31	0.15	0.60
UK	0.45	0.23	0.62	0.21	0.27	0.94
US	0.28	0.21	0.27	0.26	0.18	0.22
Total	0.58	0.20	0.37	0.29	0.24	0.50

Notes: This table shows the standardised mean deviations (SMDs) of the indicator variables (first four columns) and the resulting finance-neutral output gap estimates (last two columns) with respect to the reference model with HP-filtered indicators. See the text for the exact definition of the SMD measures. In the reference model, the cycle indicators are HP-filtered with $\lambda = 400,000$. The alternative pre-treatment methods are: de-trended indicators by second or third order polynomials; and the gap of indicator variables obtained from the semi-structural UCM.

Table 4.2: The impact of alternative pre-filtering methods on the resulting output gap estimates

	extended HP (1)	semi-struct. VAR (2)	extended HP (3)	semi-struct. VAR (4)
Reference model (*ref.*)	HP-filtered indicators	HP-filtered indicators	semi-struct. UCM	semi-struct. UCM
adjusted R2	0.66	0.01	0.41	0.15
Nb. of obs.	4101	4020	4101	4020
$\hat{u}_t - \hat{u}_t^{\text{ref.}}$	-0.24***	-0.11**	-0.76***	-0.76***
	(0.014)	*(0.044)*	*(0.027)*	*(0.042)*
$\hat{c}_t - \hat{c}_t^{\text{ref.}}$	0.46***	-0.06**	0.07***	-0.27***
	(0.014)	*(0.028)*	*(0.013)*	*(0.021)*
$\hat{cr}_t - \hat{cr}_t^{\text{ref.}}$	0.01***	-0.04***	0.04***	0.02***
	(0.002)	*(0.007)*	*(0.003)*	*(0.005)*
$\hat{hp}_t - \hat{hp}_t^{\text{ref.}}$	0.06***	0.04***	0.07***	0.07***
	(0.002)	*(0.007)*	*(0.003)*	*(0.005)*

Notes: This table presents the results from a pooled ordinary least squares estimation in which the differences in output gaps are regressed on differences in the indicator measures. The reference model is the one with HP-filtered cycle indicators (the first two columns) or the semi-structural UCM (the last two columns). Other models considered: de-trended indicators by second or third order polynomials.

A similar conclusion is reached when the output gap estimates of the semi-structural UCM are set as a reference. The same exercise as before is repeated in column 3 and 4 of Table 4.2, but the differences are calculated with respect to the semi-structural UCM. The alternative pre-treatment methods considered are: HP-filtered indicators with $\lambda = 400,000$; and de-trended indicators by second or third order polynomials. These simple linear models lead to an adjusted R2 of 41 per cent for the extended HP and 15 per cent for the semi-structural VAR. That is, the closer the cycle indicators are to the results of the semi-structural UCM, the closer the estimated gaps with the extended HP model will be to the output gap estimates of the semi-structural UCM. Crucially, the differences and similarities of these two UCMs depend on the pre-treatment method applied to the auxiliary variables. This is much less true for the semi-structural VAR.

4.4.5 Real time performance

To assess the real-time performance of the models, we compute the average revision of the output gap as $\sum_{t=t_0}^{T} \frac{1}{T-t_0+1} \frac{|\hat{y}_{it}^{t'} - \hat{y}_{it}^{t}|}{\sigma_i}$, with $\hat{y}_{it}^{t}$ and $\hat{y}_{it}^{t'}$ being the output gaps of country i at date t for a given model estimated with data respectively up to date t or $t' \in \{t+1 \text{ year}, t+3 \text{ years}\}$, σ_i the standard deviation of the baseline extended HP filter, t_0 the first date used for the exercise (2000 was chosen so that the models are stabilised for all countries) and T is the last usable date (2015 for $t' = t+1$ year or 2013 for $t' = t+3$).[13]

To meaningfully compare different models (with possibly very different output gap volatility), the revisions are normalised using the standard deviation of the most simple model, the extended HP filter in its baseline specification. This ensures that for example a 1 point revision in the output gap translates into the same standardised revision independently of the model or the set of variables included in the model. Furthermore, this normalisation neutralises the volatility of each country's cycle – FR's cycle is much less volatile than ES's, for instance – which allows comparisons between countries.

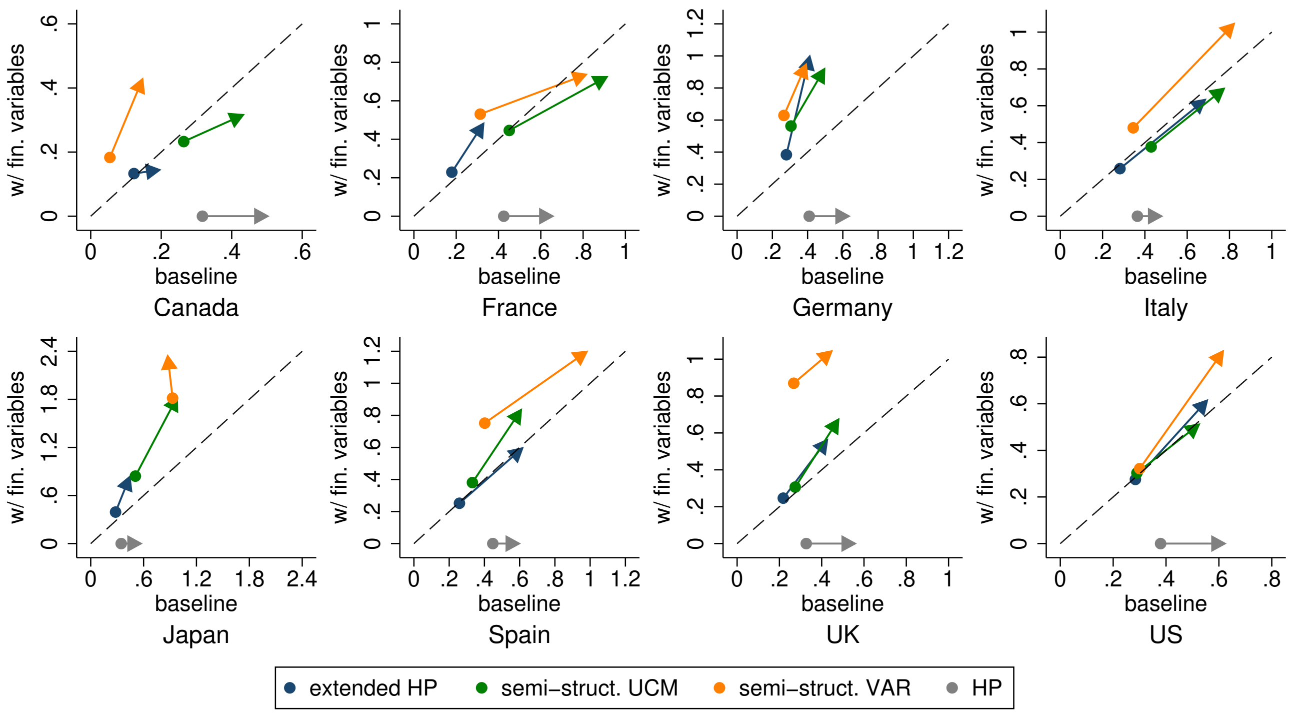
Figure 4.3 shows the standardised average errors (SAE) made at a 1-year horizon (dots at the beginning of the arrows) and at a 3-year horizon (tip of the arrows) for the baseline model ("baseline") and the model augmented with financial variables ("w/ fin. variables"). The colours of the arrows correspond to different models: the extended HP filter (blue, model A), the semi-structural UCM (green, model B) and the semi-structural VAR (orange, model C). A 45-degree dashed line is added to each plot. For comparison, the grey arrows indicate the SAE made at a 1-year horizon and at a 3-year horizon by a simple univariate HP filter (obviously, this model only has a baseline version).

Although there is no overall best performing model, the extended HP filter generally performs relatively well in terms of revision. It is the model the least revised both at the 1-year and 3-year horizon in half of the countries (FR, IT, JP and UK), independently of the presence or absence of financial variables. The extended HP filter augmented with financial variables is particularly stable compared to the other models with financial cycle indicators: it is consistently the least revised model at the 1-year horizon, and it is also the least revised after 3 years for 6 out of 8 countries.

The semi structural models are generally less stable than the reduced form model. When financial variables are added to the models, the semi-structural UCM usually outperforms the semi-structural VAR in terms of stability, while the baseline results are more mixed. With financial variables, the semi-structural VAR appears to be particularly unstable for JP, the UK, ES, and (at the 3-year horizon) for IT. For the rest, the semi-structural VAR features revisions in the range of the other models.

[13] As we use only the latest available vintage of the data, this exercise is more accurately called pseudo real-time. A real-time exercise would only use data available at the moment of the estimation. So the part of the revisions coming from the revisions to the data itself is disregarded. Yet Orphanides and Norden (2002) shows that the "the revision of published data is not the primary source of revisions in measured output gaps; the bulk of the problem is due to the pervasive unreliability of end-of-sample estimates of the trend in output."

Figure 4.3: Output gap revisions 1 and 3 years ahead



Notes: This figure shows the standardised average errors (SAE) for each model and each country made at a 1-year horizon (dots at the beginning of the arrows) and at a 3-year horizon (tip of the arrows) for the baseline model ("baseline") and the model augmented with financial variables ("w/ fin. variables"). The colours of the arrows correspond to different models: the extended HP filter (blue, model A), the semi-structural UCM (green, model B) and the semi-structural VAR (orange, model C). A 45-degree dashed line is added to each plot. If, for example, the dot (tip of the arrow) is above the 45-degree dashed line, it means that the model augmented with financial variables is less stable in real-time 1 year (3 years) ahead than the baseline model; if the arrow is inclined more than 45 degrees from the horizontal, it means that the model augmented with financial variables is more revised between 1 and 3 years ahead than the baseline model. For comparison, the grey arrows indicate the SAE made at a 1-year horizon and at a 3-year horizon by a simple (univariate) HP filter (obviously, this model has only a baseline version).

Apart from a few exceptions, financial information worsens the real-time stability of the models. Figure 4.3 shows that for almost all models and countries, the revision is higher at both the 1-year and the 3-year horizon with financial variables than in the baseline specification (the arrows are almost always above the 45-degree lines).[14]

[14]There are only two exceptions: for FR, adding financial variables to the semi-structural models seems to improve the real-time stability of the models at the 3-year horizon; and for CA even though financial cycles do not seem to play a major role, the stability of the semi-structural UCM is improved at the 3-year horizon. For the other countries, financial information worsens the real-time stability of the models or only marginally improves it.

This seems to contradict some of the previous results, in particular the conclusion of Borio et al. (2017), Grintzalis et al. (2017) and Melolinna and Tóth (2019). The reason for the discrepancy of the results is twofold. First, in the first two papers the authors use the *ex-post* gaps for each model for normalisation, i.e. they compute a revision *relative* to the model and set of variables. In contrast, the revision measure presented in this paper is *absolute,* expressed as a percentage of the standard deviation of the baseline extended HP filter. Given that the output gaps are generally more volatile with financial variables than without, the relative revision would appear to be smaller with financial variables than without even if the absolute revisions are the same. This normalisation choice matters considerably: using the standard deviation of each model and set of variables, we also find on average smaller relative revisions with financial variables than without.

Second, Borio et al. (2017) and Grintzalis et al. (2017) do not assess the impact on real-time performance of the *additional* information from financial variables to business cycle variables, but instead compare an HP filter with a model with credit and house prices only – without unemployment or capacity utilisation to account for usual business cycles. Yet Borio et al. (2014b) also shows that (*i*) the HP filter involves huge revisions, (*ii*) using unemployment alone entails the smallest revisions in the output gap, and (*iii*) adding credit and house prices growth to unemployment generates larger revisions. These conclusions are perfectly in line with our results: Figure 4.3 reveals that the baseline extended HP filter is more stable than the traditional HP filter (grey arrow). The only exception is the extended HP filter for IT, which generates relatively high revisions at the 3-year horizon.

Figure 4.4 assesses the performance of the various models before and during the Great Recession for the four countries for which financial imbalances are identified *ex-post* (Figure 4.1 shows that these countries are FR, ES, the US, and the UK). For each of these countries, we plot the real-time estimates for the different models, with or without financial variables. Each line shows the output gap as estimated using data up to date $t \in [2007, 2009, 2011, 2013, 2015]$ and on the whole sample (red line).

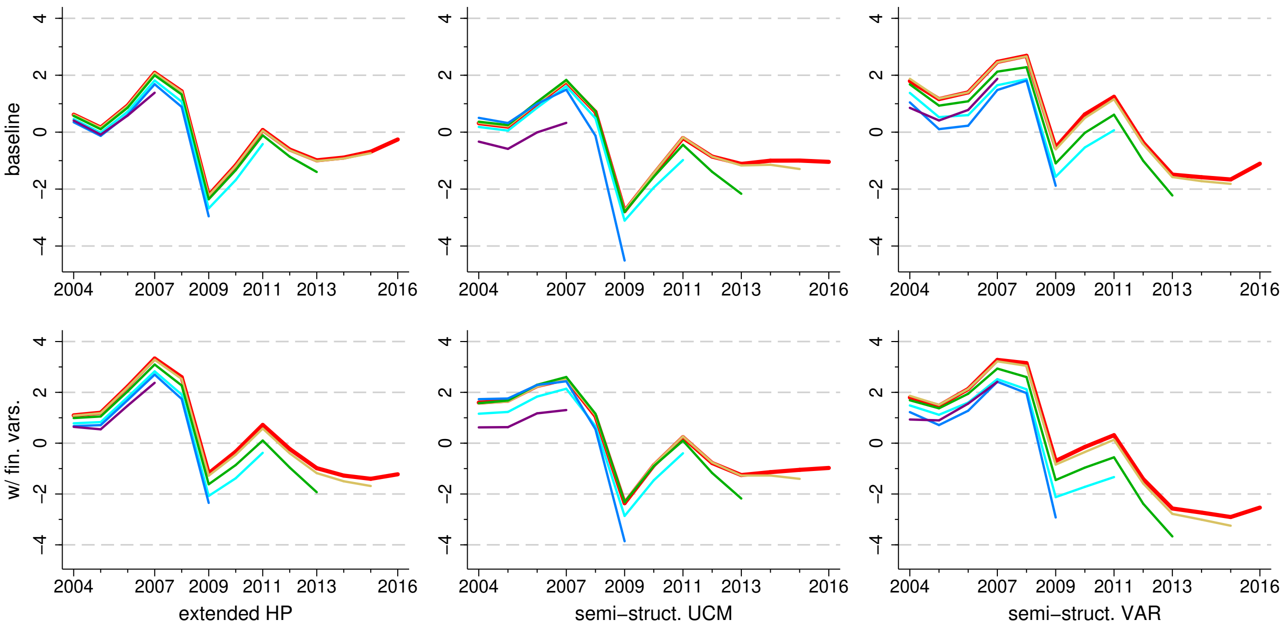
Revisions are significant for all models and countries around the crisis, at times when new information is not in line with previous developments and thus shines a new light on the present and recent past. This is equally true for models on average less revised (extended HP) and more revised (semi-structural VAR), and for models with or without financial variables.

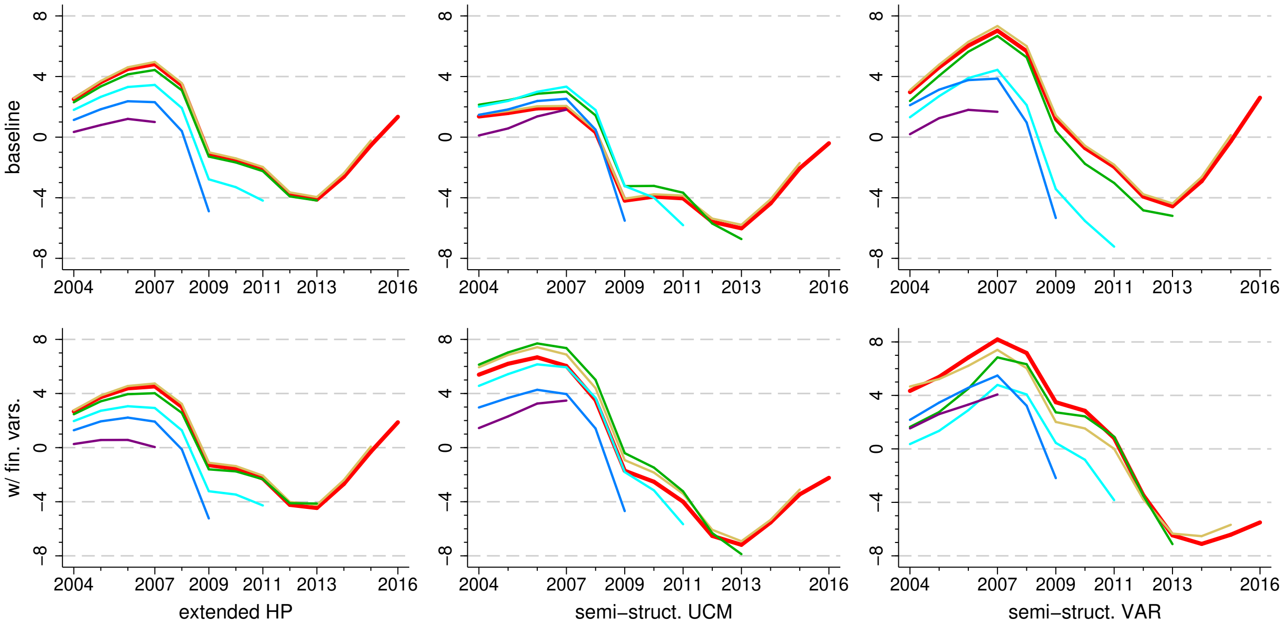
Nevertheless, the fact that the models with financial variables do not carry small revisions during the crisis does not mean that they do not carry any useful information. A robust real-time performance is an advantage, but the ability to signal (possible) unsustainable economic developments in real-time – despite the relatively large subsequent revisions – is an equally important aspect to take into consideration. As the figures show, the inclusion of financial variables consistently leads to larger output gaps as estimated before the crisis (2007, or 2006 for the US): although adding financial variables does not lead to more precise output gap estimates before or during a crisis in terms of future revisions, on average it does provide a clearer real-time signal for booms and busts. The only exception is the UK, the country for which our estimation sample starting at 1985q1 is likely to be insufficient for identifying financial cycles and capturing their impact on GDP when only the pre-crisis period is taken into account. But even in this case, the models with financial variables adjust the subsequent output gap estimates more aggressively in response to new data for the crisis period.

Finally, the figures below suggest that the semi-structural VAR model is, on average, more successful in capturing the pre-crisis overheating: for the last year before the crisis, it either indicates the largest output gap (baseline estimate for FR and US; and finance-neutral estimate for FR and ES) or similar imbalances to the other models.

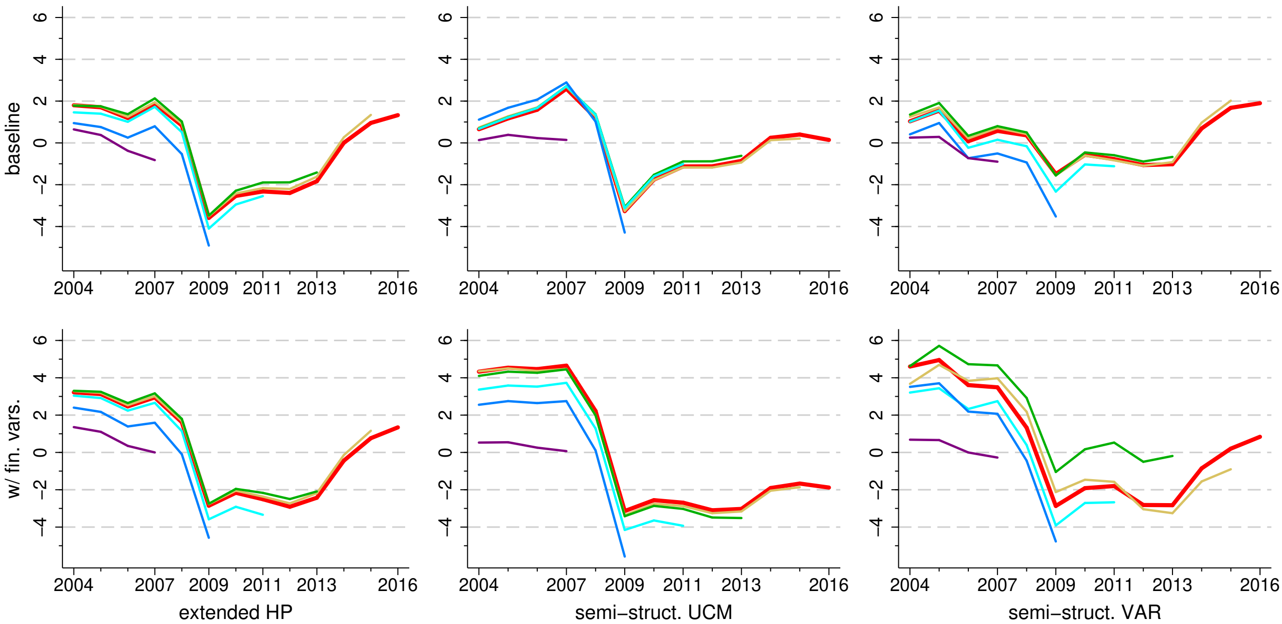
Figure 4.4: Real-time estimates

(a) France, 2007-2016

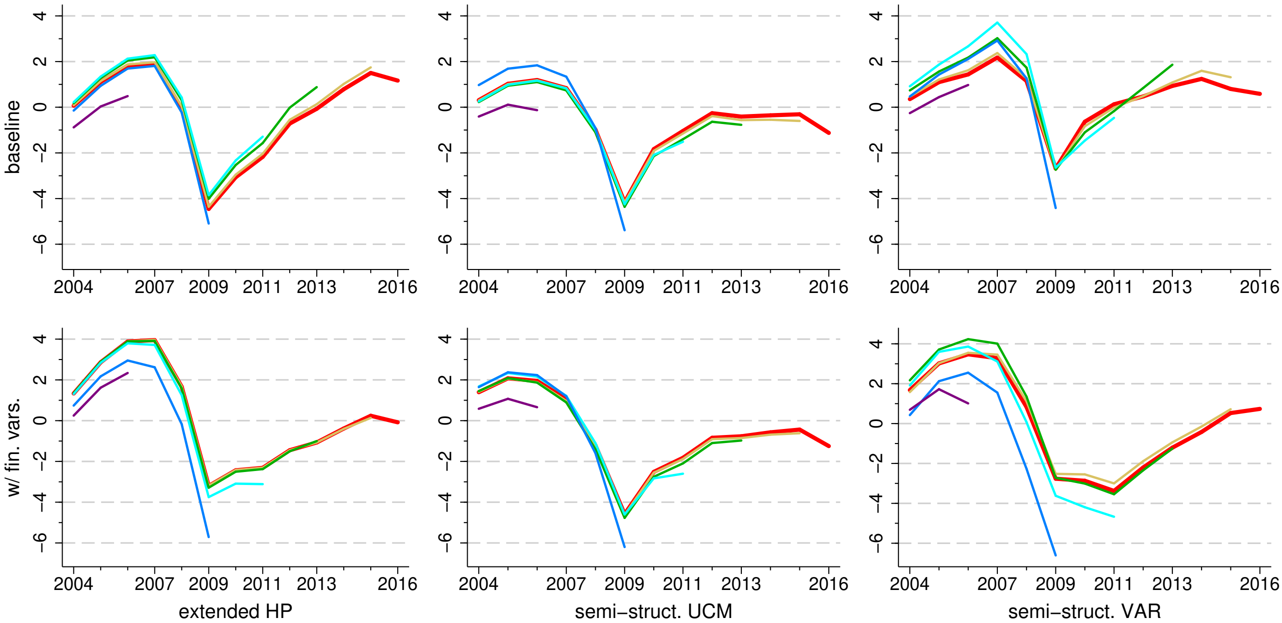




(b) Spain, 2007-2016



(c) UK, 2007-2016



(d) US, 2006-2016

Notes: Figures 4.4(a) to 4.4(d) show the real-time estimates of the extended HP filter (left), the semi-structural UCM (middle) and the semi-structural VAR (right). The graphs in the top row correspond to the baseline model and those in the bottom row to the models augmented with credit and house prices. Each line shows the output gap as estimated using data up to date $t \in [2007, 2009, 2011, 2013, 2015]$ *and on the whole sample (red line). For the US, since the crisis started earlier than in the other countries, the first estimate corresponds to 2006 instead of 2007 (purple line).*

4.4.6 Receiver operating characteristic analysis

The accuracy of the estimated output gaps and the models' early warning capabilities in real-time are assessed using receiver operating characteristic (ROC) analysis. The performance of each model (extended HP, semi-structural UCM and semi-structural VAR) and model variant (baseline and extended with financial variables) is assessed based on how accurately the estimated output gap can predict recessions.

The core of the analysis is a probit equation that relates the probability of country i being in recession ($R_{it} = 1$) to the estimated output gap h quarters earlier ($\hat{y}^{m}_{i,t-h}$) obtained from the model m. As in Section 4.4.5, the output gaps are normalised using the standard deviation of the baseline extended HP filter. Recession is defined as two consecutive quarters of negative seasonally adjusted GDP growth. We exclude "mild recessions" from the analysis, defined as a recession which lasts less than a year *and* results in a drop in GDP of less than 1 per cent. 10 such "mild" recession periods are identified. In total, we identify 33 recession periods which last between 2 and 11 quarters. The highest number of recession periods is in DE, while FR and the UK each experienced only 2 recession periods in our sample. The full list of recession periods are provided in Appendix 4.D.

By definition, R_{it} follows an autoregressive process. Therefore, the probit model also includes the lagged values of the state of the economy ($\sum_{l=1}^{L} \rho_l R_{i,t-l}$). Due to the low number of recession periods in some of the countries, no reliable and robust estimate can be obtained separately for each country. Hence, we estimate the following dynamic probit model for the panel of the eight countries:

$$\Pr\left(R_{it} = 1 | R_{i,t-1}, \ldots, R_{i,t-l}, \hat{y}_{i,t-h}\right) = \Phi\left(\sum_{l=1}^{L} \rho_l R_{i,t-l} + \delta \hat{y}^{m}_{i,t-h}\right) \tag{4.14}$$

where Φ is is the cumulative distribution function of the standard normal distribution.[15]

Using the estimated probit equation, we first predict the probability of being in recession conditional on $R_{i,t-l} = 0, l \in \{1, ..., L\}$. The predicted conditional probabilities are then used to estimate the ROC curve. For a number of different candidate probability threshold values between 0 and 1, the ROC curve plots the fraction of "true positive" events (recessions that are correctly identified by the model for a given threshold, a.k.a. *sensitivity*) against the fraction of "false positive" events (predicted recessions that did not occur, a.k.a. *1 – specificity*). The Area Under the ROC Curve (AUC) is the integral of the ROC curve, which provides a comparable aggregate metric that quantifies the overall accuracy of the prediction by taking into account both true and false signals. See Pepe (2003) for a general review of the ROC methodology.

The main difficulty of the analysis is that the performance of the estimated output gap as a leading indicator for recession cannot be clearly separated from the forecasting performance of the probit model. In particular, the optimal lag structure of eq. 4.14 may vary by country and by model. All models and countries considered, the distance between the date of entry into recession and the closest peak in the estimated output gap ranges between -1 and 29 quarters, with an average of 8.3 and a standard deviation of 6 quarters (see Appendix 4.E for the definition of our distance measure and the histogram of the distances).

Instead of relying on a single probit specification, we therefore analyse a large number of alternative probit models with different lag selection criterion. First, for each output gap estimate $\hat{y}^{m}_{it}$, the optimal lag h and the order of the autoregressive process L are selected based on the criterion that maximises the AUC.[16] Second, we re-estimate the probit equations and the corresponding AUCs using the same lag structure for all models. For example, if the optimal h in the previous exercise turn out to

[15] We estimate eq. 4.14 as a simple pooled panel probit model. A dynamic random effects probit with unobserved heterogeneity, estimated as proposed by Wooldridge (2005), yields very similar results.

[16] h is allowed to vary between 1 and 16 quarters, and L between 1 and 4 quarters. We used these boundary conditions for all other selection criteria as well. The optimal h and L are always lower than the upper bounds.

be 5, 11 and 12 for the various models, we repeat the same exercise and compare the AUCs obtained from the same probit specification for each $\hat{y}_{it}^m$ with h equal to 5, 11, and then 12. Third, we estimate a distributed lag model by replacing $\delta\hat{y}_{i,t-h}^m$ at the right-hand side of eq. 4.14 by $\sum_{j=h}^{H} \delta_j \hat{y}_{it-j}^m$. The optimal minimum ($h$) and maximum ($H$) lags are the ones that maximise the resulting AUC. Finally, instead of maximising the AUC, we used the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC) to select the optimal h and L.

The first row (a) of Table 4.3 presents the average AUCs obtained from these probit model specifications. The corresponding standard errors are calculated as $(CVC')^{1/2}$,where C is a $1 \times k$ row vector with elements $1/k$, k is the number of estimated probit specifications for a given model m, and V is the variance-covariance matrix of the estimated AUCs calculated as described in DeLong et al. (1988). To test the differences between the average AUCs, we use a $1 \times 2k$ contrast matrix $C = (1, ..., 1, -1, ..., -1)$. The first k elements with a value of 1 correspond to the vector of AUCs of a given model m, while the remaining k elements with a value of -1 correspond to the AUCs of another model to which the first one is confronted. The method for performing a test of significance for a given contrast matrix is described in detail in DeLong et al. (1988). Intuitively, the contrast matrix previously defined tests whether the sum of the AUCs obtained for the model m minus the sum of the AUCs obtained for another model is significantly different from zero. In row (b) of Table 4.3, we compare, for each model, the baseline specification with the model with financial variables. In rows (c) and (d), the AUCs obtained for the extended HP, the semi-structural UCM and the semi-structural VAR are compared, separately for the baseline model and the model with financial variables.

The second block of the table (rows e–h) presents the results for the real-time output gap estimates (see Section 4.4.5). Using the probit model estimated on the full sample, we first predict the conditional probability of being in recession with the output gap estimated on the sample up to 2000q4. For each subsequent year $t \in \{2001, ..., 2016\}$, we predict the same probability using the output gap obtained from the model estimated until t. We then compute the AUCs using the resulting real-time predictions.

Finally, the bottom panel (B) of Table 4.3 repeats the same exercise as before on the restricted sample of four countries with visible impact of financial cycles before the Great Recession: FR, ES, the UK and the US (see Section 4.4.5).

Overall, results reveal that, for each model, the average AUCs obtained from the models with financial variables are significantly higher than the corresponding AUCs of the baseline models. The only exception is the real-time forecast using the extended HP for the panel of eight countries (in row e, and the corresponding test in row f): although the model with financial variables yields higher AUC than the baseline model, the difference is not statistically significant. The difference between the AUCs is systematically higher for the restricted sample. The largest difference is obtained for the real-time forecast using the the semi-structural VAR on the restricted sample (see rows m and n).

We arrive to the same conclusion when each of the probit specification and estimation sample are separately assessed. In the vast majority of the cases, adding financial variables improves the forecasting performance of the models. Out of the 96 model comparisons (16 probit specifications $\times$ 3 models, in-sample and real-time forecasting performance), the models with financial variables yield higher AUCs in 93 cases, out of which 68 are significant at the 5% level and an additional 7 are significant at the 10% level. There are only 3 cases in which the baseline AUCs are higher than those with the financial variables, and the difference is significant in 1 case only. On the restricted sample, the superiority of the model with financial variables is even more obvious: out of the 54 comparison tests performed, the AUCs with financial variables are significantly higher at the 5% level than the baseline model in 52 cases; it is higher at the 10% significance level in 1 case; and it is higher, but the difference is not statistically significant in 1 case. Together with the tests performed on the average AUCs presented in Table 4.3, these results provide robust evidence that financial variables enhance the models' forecasting performance.

Table 4.3: Receiver operating characteristic (ROC) analysis

	extended HP		semi–struct. UCM		semi–struct. VAR	
	baseline	w/ fin. vars.	baseline	w/ fin. vars.	baseline	w/ fin. vars.
			(A) Full sample			
(a) AUC (in-sample)	0.63 (0.019)	0.65 (0.020)	0.58 (0.022)	0.69 (0.024)	0.69 (0.018)	0.73 (0.018)
(b) diff. w.r.t. baseline		0.02** (p = 0.037)		0.11*** (p < 0.001)		0.03** (p = 0.037)
(c) diff. w.r.t. extended HP			-0.05*** (p < 0.001)	0.04** (p = 0.037)	0.07*** (p < 0.001)	0.08*** (p < 0.001)
(d) diff. w.r.t. semi–struct. UCM					0.11*** (p < 0.001)	0.04* (p = 0.075)
(e) AUC (real-time)	0.58 (0.020)	0.60 (0.023)	0.61 (0.023)	0.69 (0.025)	0.60 (0.021)	0.67 (0.022)
(f) diff. w.r.t. baseline		0.02 (p = 0.253)		0.08*** (p < 0.001)		0.07*** (p = 0.004)
(g) diff. w.r.t. extended HP			0.03** (p = 0.049)	0.09*** (p < 0.001)	0.02 (p = 0.379)	0.07*** (p = 0.001)
(h) diff. w.r.t. semi–struct. UCM					-0.01 (p = 0.442)	-0.03 (p = 0.145)
			(B) Restricted sample (four countries)			
(i) AUC (in-sample)	0.68 (0.031)	0.74 (0.034)	0.66 (0.042)	0.78 (0.031)	0.72 (0.030)	0.85 (0.018)
(j) diff. w.r.t. baseline		0.06*** (p < 0.001)		0.12*** (p < 0.001)		0.13*** (p < 0.001)
(k) diff. w.r.t. extended HP			-0.03 (p = 0.139)	0.04** (p = 0.024)	0.04 (p = 0.305)	0.11*** (p = 0.001)
(l) diff. w.r.t. semi–struct. UCM					0.06 (p = 0.185)	0.07** (p = 0.018)
(m) AUC (real-time)	0.64 (0.034)	0.68 (0.037)	0.70 (0.039)	0.83 (0.023)	0.65 (0.035)	0.83 (0.018)
(n) diff. w.r.t. baseline		0.05*** (p < 0.001)		0.13*** (p < 0.001)		0.18*** (p < 0.001)
(o) diff. w.r.t. extended HP			0.06*** (p < 0.001)	0.15*** (p < 0.001)	0.02 (p = 0.502)	0.15*** (p < 0.001)
(p) diff. w.r.t. semi–struct. UCM					-0.05* (p = 0.073)	0.00 (p = 0.969)

Notes: Row (a) presents the average AUCs obtained from a set of probit model specifications. Row (b) compares, for each model, the baseline specification with the model with financial variables. In rows (c) and (d), the AUCs obtained for the various models are compared. The second block of the table (rows e–h) presents the results for the real-time output gap estimates. The bottom panel (B) shows the same results for a restricted sample of four countries: FR, ES, the UK and the US. See the text for more details.

In terms of model comparison, results are more mixed. When financial variables are added to the models, both semi-structural models seem to significantly outperform the extended HP filter (see the tests in rows c, g, k and o). Comparing the two semi-structural models with financial variables, the VAR seems to have an edge over the UCM in terms of in-sample forecasting accuracy (see rows d and l), but there is no significant difference in their forecasting performance in real-time (rows h and p). No clear message emerges for the baseline specifications.

These results are consistent with the conclusions drawn from the comparison of the results of each probit specification and estimation sample. In about 28 per cent of the cases (9 out of 32), the differences between the AUCs for the three models with financial variables are not statistically significant.[17] The semi-structural UCM has either the highest or not significantly different from the highest AUC in 21 out of the remaining 23 cases. The same number for the semi-structural UCM is 15, while there is only one case in which the extended HP shares the first place with the semi-structural VAR. To put it differently, when financial variables are included in the models, the semi-structural VAR significantly falls behind the best performing model in 2 cases only. The semi-structural UCM is the close second with 8 such cases, while the extended HP filter lags behind the best performing model in 22 cases.

At last, we also repeat the same exercise using an alternative, more flexible dependent variable in the probit equation 4.14. Instead of predicting the probability of being in recession exactly h periods ahead, R_{it} is defined as being in recession during a one-year interval between t and $t+3$. The resulting AUCs are systematically higher with this more flexible recession definition. Overall, all our previous conclusions hold true even when the more flexible probit is considered.

4.5 Discussion and conclusion

This paper estimates the sustainable growth and the finance-neutral output gap for eight advanced economies: CA, FR, DE, IT, JP, ES, the UK and the US. We test the implications of incorporating financial cycle indicators into various signal extraction models. As a first step, we estimate three "baseline" models: (A) a reduced form "extended HP filter", advanced by Borio et al. (2017), in which business and financial cycle indicators are directly incorporated into the state-space representation of the HP filter as deterministic covariates; (B) a simple semi-structural UCM featuring a Phillips curve, an Okun's law and a stochastic process relating output gap to capacity utilisation; and (C) a new approach, a multivariate semi-structural VAR with long-run restrictions *à la* Blanchard and Quah (1989) in which only one of the several structural shocks (the supply shock) is recovered. Usual business cycles are captured in these models by the unemployment rate, the capacity utilisation rate

[17] This includes the cases in which no clear winner can be identified, e.g. when A>B, A=C and B=C).

and (in the semi-structural model) CPI inflation. As a second step, the implications of adding data on credit and on house prices to capture financial cycles are tested.

The three models have different properties along several important dimensions. The comparison of the two UCMs (models A and B) assesses the practical advantages and the potential pitfalls of treating the cycle indicator variables as deterministic (as in model A) compared to modelling them as stochastic processes (as in model B). On the one hand, the reduced form extended HP filter is particularly appealing for its simplicity and real-time stability. Indeed, there are only a limited number of parameters to estimate, the identification of the model is quite straightforward and, as a consequence, the model exhibits relatively stable output gap estimates in terms of real-time revisions. With financial variables, this model is particularly stable compared to the other models: it is the least revised model at the 1-year horizon for all countries considered, and it is also the least revised after 3 years for 6 out of 8 countries.

On the other hand, this approach requires accurate and readily available cycle indicators to achieve unbiased output gap estimates. Since "pure" business or financial cycle indicators unaffected by structural changes are in practice almost never available, some form of pre-transformation should be applied to the auxiliary variables before feeding them into the model. A variety of alternative stationary-inducing transformations can be employed. For example, if the time series is supposed to be trend stationary, de-trending the data by regressing out a – linear or higher order, depending on properties of the time series – time trend is a valid approach. A different procedure involves detrending the data by filtering out very low frequency signals, but still allowing the trend to slowly evolve following a stochastic process. In principle, any pre-transformation is valid if it leads to stationary time series. The difficulty is to select the best approach that results in the most accurate cycle indicator.[18]

Results suggest that the choice of the pre-treatment method is very important. If the pre-transformed indicator does not accurately capture the cyclical pattern of the variable at the relevant frequency range, the variable plugged into the model can either receive a positive weight and transmit the bias to the estimated output gap, or with a (close to) zero coefficient it cannot impact the estimated output gap. For example, our results indicate that a relatively standard pre-treatment of the business and financial cycle indicators using a HP filter with smoothing parameter (λ) of 400,000 leads to a likely underestimation of the importance of credit cycles in the UK. In fact, this approach can be viewed as a two step procedure, in which possible structural shocks to the indicator variables are first removed before their impact on the cyclical pattern of GDP is assessed. By re-estimating the model for all countries using various alternative trend removal techniques, we show that close to half of the differences in the output gap estimates compared to the semi-structural UCM can be explained by the differences

[18] In a broader sense, the seasonal adjustment applied to original series is also a form of pre-transformation. While the seasonal adjustment filters out signals at the yearly frequency, de-trending the series (either by regressing out a deterministic trend or by applying a filter) removes very low frequency signals.

between the pre-treated cycle indicators and the cyclical parts of the same indicator estimated using the semi-structural UCM.

Since the semi-structural UCM allows for stochastic shocks to the trend of business cycle indicators and financial variables, the model can be viewed as the simultaneous estimation of these two steps. However, this flexibility is not without a price. Although the method does not require pre-transforming the business and financial cycle indicators, it does require introducing potentially strong prior judgements by choosing the structure of the model and by defining Bayesian prior distributions on the parameters. To illustrate the benefits and potential shortcomings of this approach, this paper presents a simple yet widely used UCM with very light structure and standard Bayesian prior beliefs about the parameters. While the baseline output gap estimates are generally close to those obtained using the extended HP filter, there are several examples showing that the two UCMs use the additional information provided by the financial variables differently. For example, the cases of JP and – to a lesser extent – the US illustrate how different approaches trigger different results and give a warning that relying on a more flexible approach may also lead to misleading conclusions. More in line with the Japanese "Lost Decade" paradigm, the extended HP filter reveals a significantly higher and more prolonged impact of the Japanese financial bubble of the 1990s than the semi-structural UCM. Similarly, the reduced form model seems to better able to signal the "overheating" of the US economy before the global financial crisis than the semi-structural UCM.

The proposed semi-structural VAR addresses the difficulty of estimating the unobserved output gap from a different angle. Similarly to the bivariate Blanchard and Quah (1989) model, the beauty of this approach is that the identifying assumption perfectly matches with the definition of potential growth: only the (supply) shocks to the trend have permanent effects on GDP. In contrast to the original model, we exploit a wider set of information by using several (both regular business and financial) cycle indicators without imposing further restrictions. Since the model uses different identifying assumptions to recover the output gap, it can provide very informative insights in addition to the other UCM-based models – especially so since the two UCMs share the same basic structure. In the same way, it is also an ideal candidate for external validation of the UCMs' results. For example, the model reinforces the findings from the extended HP filter for JP and the US, while it rather gives credit to the semi-structural UCM in the case of the UK. As with the extended HP filter, one drawback of the semi-structural VAR is that the model is sensitive to the pre-treatment method applied to the auxiliary variables. However, compared to the reduced form UCM approach, the change in the resulting output gap estimates is much less predictable.

Both in terms of model used or the sets of variables incorporated in the models, our results highlight a possible trade-off between accurately disentangling short-term and long-term (sustainable) economic dynamics and the real-time stability of the output gap estimates. The deterioration of the real-time stability of the model with financial variables (compared to the model which already includes relevant business cycle indicators, such as unemployment and capacity utilisation) is a common feature of all approaches. The increased (absolute) revisions is typical both in the short-run and the longer-run, sometimes even for countries without clear financial cycles identified ex-post (most importantly for DE).

Although the stability of a model is an important aspect – since a model substantially revised when new data becomes available is of limited use for policy analysis –, relatively large revisions do not necessarily preclude the usefulness of the model.[19] On the contrary: the models with financial variables included generally prove to be more successful in identifying unsustainable economic developments both ex-post and in real-time. By assessing the three models' forecasting performance, the paper shows that including financial variables significantly improves the models' accuracy in predicting recessions. This is also true when real-time output gap estimates are used, in particular for the countries in which large pre-crisis imbalances are identified ex-post. For these countries, the models with financial variables consistently indicate larger pre-crisis (positive) output gap already before the outbreak of the crisis. To put it into policy perspective, credit and house prices in ES, for example, would have signalled a much larger boom before the 2008-09 crisis and could have helped the authorities to mitigate its impact.

In a similar vein, the trade-off between accuracy and real-time stability is also apparent when the various models are compared. While the semi structural models are generally less stable than the reduced form model, especially when financial variables are added to the models, they also seem to better capture the increase in macroeconomic and financial imbalances and to more accurately predict future recessions. Our proposed semi-structural VAR augmented with financial variables performs particularly well in terms of predicting power: it features the best in-sample forecasting performance of recession probabilities among all three models, and it has similar real-time prediction capabilities to the semi-structural UCM and clearly superior to the extended HP filter.

This brings us to the following more general conclusions. First, financial information should be taken into account when assessing the cyclical position of the economy. For most countries, there is already enough data and variability to estimate the interaction between financial cycles and GDP. Obviously, there is no guarantee that the models augmented with financial variables would accurately capture the build-up of imbalances coming from other sources. Moreover, even if the source is identified, it may not be possible to empirically assess its impact on GDP if the sample does not cover a sufficient number of cycles for this source. Nevertheless, the real-time results for the UK suggest that it is still beneficial to incorporate relevant information into the model *after* the burst of the bubble: the model may be quicker to adjust the subsequent output gap estimates as new data points become available.

Second, given the large uncertainty surrounding output gap estimates, as also illustrated by our results, relying on a single model (or a few closely related models that can be seen as model variants)

[19] To give an extreme example, an unrealistic rigid model such as $\hat{y}_t = \pi_t - 2$ is admittedly never revised, yet its poor accuracy makes it irrelevant.

is hardly enough to draw robust conclusions. In this regard, the proposed semi-structural VAR is an ideal complement of the more widely used UCM approaches. This paper does not propose a statistical method for weighting the different results. Instead, in order to avoid policy mistakes caused by, for example, false and short-lived signals of the more flexible approaches or, inversely, by failing to consider relevant information hidden (smoothed out) by the more stable models, it is sensible to monitor the results obtained from several methods.

Obviously, this paper does not cover all possible methods, not even within the UCM-based statistical filtering family. The comparison of the two UCMs presented in the paper illustrate one particularly relevant aspect of the methods, namely the implications of treating the (financial) cycle indicator variables as deterministic compared to modelling them as stochastic processes. Moreover, results from an alternative reduced form approach (model A′, the "dynamic multivariate filter", presented in Appendix 4.B) suggest that the exact formulation of the reduced form model is of limited practical relevance. Yet, the possibilities of adjusting the semi-structural UCM by imposing more structure on it are endless. Furthermore, by imposing the full general equilibrium structure of the model, we arrive at a second class of methods: the structural approaches (see e.g. the classification of Mishkin (2007)). In this approach, potential output is the "frictionless" part of GDP. Finally, according to the classification of Mishkin (2007), a separate class of methods is dedicated to the "production function" approach. However, from the methodological point of view this approach is not a trend-cycle decomposition technique *per se:* the input factors need to be decomposed – individually or simultaneously – before recovering the potential growth and the output gap. These trend-cycle decompositions are usually carried out using simple univariate or small multivariate UCMs, but a semi-structural model with an additional restriction given by the production function relating the different (trend) input factors to (trend) GDP can, in principle, also be considered. Whether or not these additional restrictions are helpful to recover the finance-neutral output gap is ultimately an empirical question. We leave this for future research.

Bibliography

Alberola-Ila, Enrique, Rocío Gondo, Marco Jacopo Lombardi, and Diego Urbina, "Output gaps and policy stabilisation in Latin America: the effect of commodity and capital flow cycles," Technical Report 568, Bank for International Settlements June 2016.

Amador-Torres, Juan S., Jose Eduardo Gomez-Gonzalez, Jair N. Ojeda-Joya, Oscar F. Jaulin-Mendez, and Fernando Tenjo-Galarza, "Mind the gap: Computing finance-neutral output gaps in Latin-American economies," *Economic Systems*, September 2016, *40* (3), 444–452.

Amisano, Gianni and Carlo Giannini, *Topics in Structural VAR Econometrics*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1997.

Anderton, Robert, Ted Aranki, Alistair Dieppe, Catherine Elding, Stephan Haroutunian, Pascal Jacquinot, Valerie Jarvis, Vincent Labhard, Desislava Rusinova, and Béla Szörfi, "Potential output from a euro area perspective," Occasional Paper Series 156, European Central Bank 2014.

Andrle, Michal, "What Is in Your Output Gap? Unified Framework & Decomposition into Observables," IMF Working Paper 13/105, International Monetary Fund 2013.

Anundsen, André K., Karsten Gerdrup, Frank Hansen, and Kasper Kragh-Sørensen, "Bubbles and Crises: The Role of House Prices and Credit," *Journal of Applied Econometrics*, 2016, *31* (7), 1291–1311.

Anvari, Vafa, Neléne Ehlers, and Rudi Steinbach, "A semi-structural approach to estimate South Africa's potential output," Technical Report 6504, South African Reserve Bank November 2014.

Bauer, Gregory H. and Eleonora Granziera, "Monetary Policy, Private Debt, and Financial Stability Risks," *International Journal of Central Banking*, 2017, *13* (3), 337–373.

Bayoumi, Tamim, "The morning after: explaining the slowdown in Japanese growth in the 1990s," *Journal of International Economics*, April 2001, *53* (2), 241–259.

Benes, Jaromir, Kevin Clinton, Roberto Garcia-Saltos, Marianne Johnson, Douglas Laxton, Petar Manchev, and Thornton Matheson, "Estimating Potential Output with a Multivariate Filter," IMF Working Paper 10/285, International Monetary Fund 2010.

Berger, Helge, Thomas Dowling, Sergi Lanau, Mico Mrkaic, Pau Rabanal, and Marzie Taheri Sanjani, "Steady as She Goes—Estimating Potential Output During Financial "Booms and Busts"," Technical Report 15/233, International Monetary Fund November 2015.

Bernanke, Ben S. and Ilian Mihov, "Measuring Monetary Policy," *The Quarterly Journal of Economics*, August 1998, *113* (3), 869–902.

_ , Mark Gertler, and Simon Gilchrist, "The financial accelerator in a quantitative business cycle framework," in "Handbook of Macroeconomics," Vol. 1, Elsevier, January 1999, pp. 1341–1393.

Bernhofer, Dominik, Octavio Fernández-Amador, Martin Gächter, and Friedrich Sindermann, "Finance, Potential Output and the Business Cycle: Empirical Evidence from Selected Advanced and CESEE Economies," *Focus on European Economic Integration*, 2014, (2), 52–75.

Blanchard, Olivier Jean and Danny Quah, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, September 1989, *79* (4), 655–673.

Bordo, Michael D., Angela Redish, and Hugh Rockoff, "Why didn't Canada have a banking crisis in 2008 (or in 1930, or 1907, or ...)?," *The Economic History Review*, 2015, *68* (1), 218–243.

Borio, Claudio, "The financial cycle and macroeconomics: What have we learnt?," *Journal of Banking & Finance*, August 2014, *45*, 182–198.

_ **and Philip Lowe**, "Asset prices, financial and monetary stability: exploring the nexus," Technical Report 114, Bank for International Settlements July 2002.

_ **, Mathias Drehmann, and Kostas Tsatsaronis**, "Stress-testing macro stress testing: Does it live up to expectations?," *Journal of Financial Stability*, June 2014, *12*, 3–15.

_ **, Piti Disyatat, and Mikael Juselius**, "A parsimonious approach to incorporating economic information in measures of potential output," BIS Working Paper 442, Bank for International Settlements 2014.

_ **,** _ **, and** _ , "Rethinking potential output: embedding information about the financial cycle," *Oxford Economic Papers*, July 2017, *69* (3), 655–677.

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones, "How do business and financial cycles interact?," *Journal of International Economics*, May 2012, *87* (1), 178–190.

Coibion, Olivier, Yuriy Gorodnichenko, and Mauricio Ulate, "The Cyclical Sensitivity in Estimates of Potential Output," *Brookings Papers on Economic Activity*, 2018, *49* (2 (Fall)), 343–441. Publisher: Economic Studies Program, The Brookings Institution.

Committee, Besel, "Guidance for national authorities operating the countercyclical capital buffer," Technical Report, Basel, Switzerland December 2010.

DeJong, David N. and Charles H. Whiteman, "Estimating Moving Average Parameters: Classical Pileups and Bayesian Posteriors," *Journal of Business & Economic Statistics*, July 1993, *11* (3), 311–317.

DeLong, Elizabeth R., David M. DeLong, and Daniel L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, 1988, *44* (3), 837–845.

Detken, Carsten, Tuomas A. Peltonen, Willem Schudel, and Markus Behn, "Setting countercyclical capital buffers based on early warning models: would it work?," Technical Report 1604, European Central Bank November 2013.

Edge, Rochelle M. and Ralf R. Meisenzahl, "The Unreliability of Credit-to-GDP Ratio Gaps in Real Time: Implications for Countercyclical Capital Buffers," *International Journal of Central Banking*, December 2011, *7* (4), 261–298.

Ermolaev, Igor, Michel Juillard, Ioan Carabenciov, Charles Freedman, Douglas Laxton, Ondrej Kamenik, and Dmitry Korshunov, "A Small Quarterly Projection Model of the US Economy," IMF Working Paper 08/278, International Monetary Fund 2008.

Felipe, Jesus, Noli Sotocinal, and Connie Bayudan-Dacuycuy, "The Impact of Financial Factors on the Output Gap and Estimates of Potential Output Growth," Technical Report 457, Asian Development Bank October 2015.

Gali, Jordi, "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review*, March 1999, *89* (1), 249–271.

Grant, Angelia L. and Joshua C. C. Chan, "Reconciling output gaps: Unobserved components model and Hodrick–Prescott filter," *Journal of Economic Dynamics and Control*, February 2017, *75*, 114–121.

Grintzalis, Ioannis, David Lodge, and Ana-Simona Manu, "The implications of global and domestic credit cycles for emerging market economies: measures of finance-adjusted output gaps," Technical Report 2034, European Central Bank March 2017.

Haltom, Renee Courtois, "Why Was Canada Exempt from the Financial Crisis?," *Econ Focus*, 2013, (4Q), 22–25.

Harvey, A. C. and A. Jaeger, "Detrending, stylized facts and the business cycle," *Journal of Applied Econometrics*, July 1993, *8* (3), 231–247.

Hume, Michael and Andrew Sentance, "The global credit boom: Challenges for macroeconomics and policy," *Journal of International Money and Finance*, December 2009, *28* (8), 1426–1461.

Jong, Piet De, "The Diffuse Kalman Filter," *The Annals of Statistics*, 1991, *19* (2), 1073–1083.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor, "When Credit Bites Back," *Journal of Money, Credit and Banking*, December 2013, *45* (s2), 3–28.

Kilian, Lutz and Helmut Lütkepohl, *Structural Vector Autoregressive Analysis*, Cambridge New York, NY: Cambridge University Press, November 2017.

Kim, Chang-Jin and Jaeho Kim, "The 'Pile-up Problem' in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives," Technical Report 51118, University Library of Munich, Germany October 2013.

King, Robert G. and Sergio T. Rebelo, "Low frequency filtering and real business cycles," *Journal of Economic Dynamics and Control*, January 1993, *17* (1), 207–231.

_ , Charles I. Plosser, James H. Stock, and Mark W. Watson, "Stochastic Trends and Economic Fluctuations," *The American Economic Review*, 1991, *81* (4), 819–840.

Kiyotaki, Nobuhiro and John Moore, "Credit Cycles," *Journal of Political Economy*, April 1997, *105* (2), 211–248.

Krupkina, Anna, Elena Deryugina, and Alexey Ponomarenko, "Estimating Sustainable Output Growth in Emerging Market Economies," *Comparative Economic Studies*, March 2015, *57* (1), 168–182.

Kuttner, Kenneth N., "Estimating Potential Output as a Latent Variable," *Journal of Business & Economic Statistics*, July 1994, *12* (3), 361–368.

Lang, Jan Hannes, Cosimo Izzo, Stephan Fahr, and Josef Ruzicka, "Anticipating the bust: a new cyclical systemic risk indicator to assess the likelihood and severity of financial crises," Technical Report 219, European Central Bank February 2019. Occasional Paper Series.

Maliszewski, Wojciech and Longmei Zhang, "China's Growth; Can Goldilocks Outgrow Bears?," Technical Report 15/113, International Monetary Fund May 2015.

Melolinna, Marko and Máté Tóth, "Output gaps, inflation and financial cycles in the UK," *Empirical Economics*, March 2019, *56* (3), 1039–1070.

Mishkin, Frederic S., "Estimating potential output: a speech at the Conference on Price Measurement for Monetary Policy, Federal Reserve Bank of Dallas, Dallas, Texas, May 24, 2007," Technical Report 299, Board of Governors of the Federal Reserve System (U.S.) 2007.

Odor, Ludovit and Judita Jurasekova Kucserova, "Finding Yeti: More robust estimates of output gap in Slovakia," Technical Report WP 1/2014, Research Department, National Bank of Slovakia March 2014.

Okun, Arthur Melvin, "Potential GNP & Its Measurement and Significance," *American Statistical Association, Proceedings of the Business and Economics Statistics Section*, 1962, pp. 98–104.

Orphanides, Athanasios and Simon van Norden, "The Unreliability of Output-Gap Estimates in Real Time," *The Review of Economics and Statistics*, November 2002, *84* (4), 569–583.

Pepe, Margaret Sullivan, *The Statistical Evaluation of Medical Tests for Classification and Prediction* Oxford Statistical Science Series, Oxford, New York: Oxford University Press, 2003.

Rosenberg, Barr, "Random Coefficients Models: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression," in "NBER Chapters," National Bureau of Economic Research, Inc, 1973, pp. 399–428.

Rünstler, Gerhard and Marente Vlekke, "Business, housing, and credit cycles," *Journal of Applied Econometrics*, March 2018, *33* (2), 212–226.

Schularick, Moritz and Alan M. Taylor, "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, April 2012, *102* (2), 1029–1061.

Shephard, Neil, "Maximum Likelihood Estimation of Regression Models with Stochastic Trend Components," *Journal of the American Statistical Association*, June 1993, *88* (422), 590–595.

Stock, James H., "Unit roots, structural breaks and trends," in "Handbook of Econometrics," Vol. 4, Elsevier, January 1994, pp. 2739–2841.

_ **and Mark W. Watson**, "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, March 1998, *93* (441), 349–358.

White, William R., "Is price stability enough?," Technical Report 205, Bank for International Settlements April 2006.

Woodford, Michael, *Interest and prices : foundations of a theory of monetary policy / Michael Woodford*, Princeton, N.J. : Woodstock: Princeton University Press, 2003. Includes bibliographical references (p. 747-764) and index.

Wooldridge, Jeffrey M., "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 2005, *20* (1), 39–54.

APPENDIX

4.A Data sources

Table 4.4: Data sources

Variable	Main sources	Other sources
Total Gross Domestic Product in Constant Prices	OECD, "Main Economic Indicators - complete database", Main Economic Indicators (database), http://dx.doi.org/10.1787/data-00052-en	For Canada, Germany, Spain, Italy and Japan: GDP is computed with GDP growth rate from the OECD (Sources : Leading Indicators OECD: Reference Series: Gross Domestic Product). For the United States and the United Kingdom, GDP is in chained 2000 National Currency Units.
Consumer Price Index of All Items ; Index 2010	OECD, "Main Economic Indicators - complete database", Main Economic Indicators (database), http://dx.doi.org/10.1787/data-00052-en	
Rate of Capacity utilisation	OECD, "Main Economic Indicators - complete database", Main Economic Indicators (database),http://dx.doi.org/10.1787/data-00052-en	For Canada, Industrial Capacity utilisation rates are from Statistics Canada. For Japan, data are "Operating Ratio index" from "Indices of Industrial Production, Ministry of Economy, Trade and Industry, Statistics".
Unemployment Rate ; Aged 15-74: All Persons	OECD Economic Outlook database	For Germany, data are from OECD, "Main Economic Indicators - complete database", Main Economic Indicators (database),http://dx.doi.org/10.1787/data-00052-en
Total credit to private non-financial sector, all sectors, adjusted for breaks	BIS total credit statistics	
Real Residential Property Prices; Long Series; Index 1995	National sources, BIS Residential Property Price database	

4.B Dynamic multivariate filter

The model is as follows:

$$\begin{cases} \hat{y}_t = \alpha_1 \hat{y}_{t-1} + \gamma' z_t + \varepsilon_t^{\hat{y}} \\ y_t^\star = y_{t-1}^\star + g_t^y + \varepsilon_t^{y^\star} \\ g_t^y = \alpha_2 \bar{g}^y + (1-\alpha_2)\, g_{t-1}^y + \varepsilon_t^{gy} \end{cases} \tag{4.15}$$

where the notations are similar as in model A (section 4.2.1). $\bar{g}^y$ is calibrated as the sample average growth rate. $\varepsilon_t^{\hat{y}}$, $\varepsilon_t^{y^\star}$ and ε_t^{gy} are i.i.d. error terms. To achieve stationarity, we apply the same pre-transformation technique on the additional variables z_t as in the extended HP filter.

This model is arguably more flexible than the extended HP filter, but the increased number of parameters to be estimated makes the identification less straightforward. In particular, since the scaling factor is now estimated instead of calibrated, the "pile-up problem" becomes a major issue. To overcome the difficulties in estimating the model, we use Bayesian techniques (see DeJong and Whiteman (1993); Kim and Kim (2013)). The priors and the posterior estimates are presented in Table 4.5. Results are shown in Figure 4.5.

Figure 4.5: Output gap estimates: dynamic multivariate filter

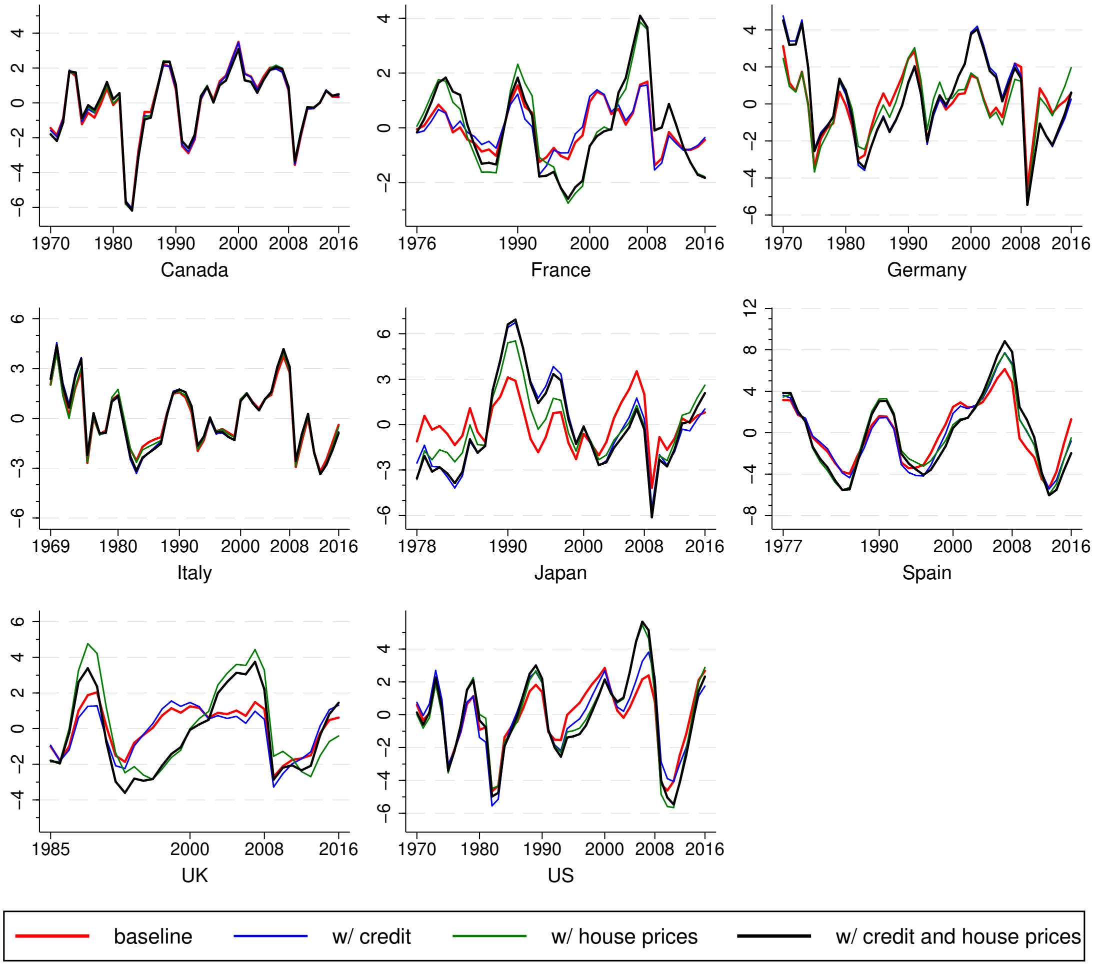


Table 4.5: Parameter estimates: dynamic multivariate filter (part 1)

	model	α_1		α_2		$\bar{g}^y$		u		c	
Prior density type		beta		gamma		beta		gamma		gamma	
Prior	(a)	0.6	*(0.2)*	0.3	*(0.1)*	$\overline{\Delta y_t}$*	*(0.001)*	0.5	*(0.3)*	0.5	*(0.3)*
	(b)	0.6	*(0.2)*	0.3	*(0.1)*	$\overline{\Delta y_t}$*	*(0.001)*	0.5	*(0.3)*	0.5	*(0.3)*
Posterior											
CA	(a)	0.3373	*(0.0383)*	0.0560	*(0.0008)*	0.0067	*(0.0664)*	0.5334	*(0.1148)*	0.2236	*(0.0263)*
	(b)	0.3418	*(0.0380)*	0.0592	*(0.0008)*	0.0067	*(0.0677)*	0.4757	*(0.1267)*	0.2228	*(0.0264)*
FR	(a)	0.6748	*(0.0613)*	0.1249	*(0.0007)*	0.0048	*(0.1286)*	0.1691	*(0.0988)*	0.0807	*(0.0316)*
	(b)	0.6107	*(0.0649)*	0.1247	*(0.0006)*	0.0048	*(0.1083)*	0.1838	*(0.0922)*	0.0762	*(0.0276)*
DE	(a)	0.2793	*(0.0139)*	0.0353	*(0.0009)*	0.0051	*(0.0837)*	0.3763	*(0.1628)*	0.2850	*(0.0384)*
	(b)	0.2689	*(0.0437)*	0.0702	*(0.0008)*	0.0050	*(0.0732)*	0.3038	*(0.1524)*	0.2886	*(0.0349)*
IT	(a)	0.4929	*(0.0527)*	0.0373	*(0.0009)*	0.0045	*(0.0784)*	0.3667	*(0.1188)*	0.2728	*(0.0371)*
	(b)	0.4858	*(0.0120)*	0.0125	*(0.0009)*	0.0044	*(0.0703)*	0.3352	*(0.1183)*	0.2837	*(0.0362)*
JP	(a)	0.4406	*(0.0141)*	0.0177	*(0.0009)*	0.0052	*(0.0835)*	0.3667	*(0.1989)*	0.1265	*(0.0182)*
	(b)	0.3781	*(0.0076)*	0.0096	*(0.0009)*	0.0052	*(0.0767)*	0.3842	*(0.2005)*	0.1308	*(0.0169)*
ES	(a)	0.4947	*(0.0280)*	0.0431	*(0.0008)*	0.0052	*(0.1099)*	0.3667	*(0.0809)*	0.0647	*(0.0260)*
	(b)	0.4168	*(0.0325)*	0.0487	*(0.0008)*	0.0053	*(0.0901)*	0.2770	*(0.0761)*	0.0810	*(0.0273)*
UK	(a)	0.6606	*(0.0615)*	0.1234	*(0.0008)*	0.0056	*(0.1141)*	0.2758	*(0.1471)*	0.0846	*(0.0309)*
	(b)	0.6245	*(0.0517)*	0.0892	*(0.0007)*	0.0054	*(0.0967)*	0.1712	*(0.1025)*	0.0641	*(0.0273)*
US	(a)	0.2552	*(0.0833)*	0.2059	*(0.0006)*	0.0068	*(0.0873)*	1.3852	*(0.2436)*	0.2586	*(0.1557)*
	(b)	0.2393	*(0.0793)*	0.1882	*(0.0007)*	0.0067	*(0.0802)*	1.2834	*(0.2445)*	0.2817	*(0.1558)*

*Notes: This table shows the estimated parameters for the dynamic multivariate filter (model A') and their standard deviation (in parentheses). Rows (a) show the estimations of the baseline model, while rows (b) present the estimations of the model augmented with credit and house prices. * $\overline{\Delta y_t}$ is the country-level average GDP growth rate.*

Table 4.6: Parameter estimates: dynamic multivariate filter (part 2)

	model	cr		hp		$\varepsilon_t^{y^\star}$		$\varepsilon_t^{\hat{y}}$		ε_t^{gy}	
Prior density type		normal		normal		invgamma		invgamma		invgamma	
Prior	(a)					0.001	*(∞)*	0.2*	*(∞)*	0.001	*(∞)*
	(b)	0.5	*(0.3)*	0.5	*(0.3)*	0.001	*(∞)*	0.2*	*(∞)*	0.001	*(∞)*
Posterior											
CA	(a)					0.0007	*(0.0003)*	0.0050	*(0.0006)*	0.0009	*(0.0022)*
	(b)	0.0138	*(0.0341)*	0.0147	*(0.0165)*	0.0008	*(0.0022)*	0.0050	*(0.0027)*	0.0010	*(0.0050)*
FR	(a)					0.0005	*(0.0004)*	0.0040	*(0.0005)*	0.0014	*(0.0014)*
	(b)	-0.0410	*(0.0357)*	0.0510	*(0.0216)*	0.0005	*(0.0014)*	0.0040	*(0.0022)*	0.0013	*(0.0059)*
DE	(a)					0.007	*(0.0002)*	0.0069	*(0.0003)*	0.0012	*(0.0020)*
	(b)	0.2097	*(0.0751)*	0.0215	*(0.0598)*	0.0006	*(0.0020)*	0.0068	*(0.0027)*	0.0012	*(0.0023)*
IT	(a)					0.0009	*(0.0007)*	0.0072	*(0.0007)*	0.0009	*(0.0016)*
	(b)	0.0160	*(0.0226)*	0.0071	*(0.0150)*	0.0010	*(0.0017)*	0.0073	*(0.0056)*	0.0006	*(0.0051)*
JP	(a)					0.0006	*(0.0002)*	0.0088	*(0.0003)*	0.0007	*(0.0039)*
	(b)	0.1017	*(0.0508)*	0.0728	*(0.0432)*	0.0007	*(0.0039)*	0.0085	*(0.0050)*	0.0005	*(0.0046)*
ES	(a)					0.0006	*(0.0002)*	0.0055	*(0.0021)*	0.0010	*(0.0015)*
	(b)	0.0301	*(0.0240)*	0.0486	*(0.0143)*	0.0006	*(0.0015)*	0.0053	*(0.0089)*	0.0007	*(0.0110)*
UK	(a)					0.0006	*(0.0005)*	0.0052	*(0.0009)*	0.0018	*(0.0019)*
	(b)	-0.0534	*(0.0265)*	0.0681	*(0.0166)*	0.0006	*(0.0018)*	0.0052	*(0.0068)*	0.0011	*(0.0098)*
US	(a)					0.0007	*(0.0005)*	0.0052	*(0.0007)*	0.0022	*(0.0007)*
	(b)	0.0478	*(0.0603)*	0.0731	*(0.0325)*	0.0006	*(0.0007)*	0.0052	*(0.0035)*	0.0022	*(0.0057)*

*Notes: see notes from Table 4.5. * 0.4 for JP*

4.C Parameter estimates

For reasons of brevity reasons, only the estimated parameters of the baseline (model a) and the most complete model with both credit and house prices included (model b) are shown. The other estimation results are available from the authors upon request.

Table 4.7: Parameter estimates: extended Hodrick-Prescott filter

	model	u		c		cr		hp	
CA	(a)	-1.01	*(0.10)*	0.24	*(0.02)*				
	(b)	-0.97	*(0.11)*	0.25	*(0.02)*	0.04	*(0.03)*	-0.01	*0.02)*
FR	(a)	-0.63	*(0.14)*	0.26	*(0.03)*				
	(b)	-0.48	*(0.16)*	0.21	*(0.03)*	-0.01	*(0.04)*	0.08	*(0.02)*
DE	(a)	-1.17	*(0.20)*	0.32	*(0.03)*				
	(b)	-0.51	*(0.21)*	0.38	*(0.03)*	0.24	*(0.07)*	0.06	*(0.05)*
IT	(a)	-0.89	*(0.16)*	0.45	*(0.03)*				
	(b)	-0.66	*(0.18)*	0.47	(0.03)	0.01	*(0.03)*	0.05	*(0.02)*
JP	(a)	-1.35	*(0.41)*	0.16	*(0.02)*				
	(b)	-0.99	*(0.43)*	0.15	*(0.02)*	0.03	*(0.08)*	0.22	*(0.06)*
ES	(a)	-0.60	*(0.09)*	0.06	*(0.03)*				
	(b)	-0.53	*(0.10)*	0.06	*(0.03)*	-0.05	*(0.04)*	0.05	*(0.02)*
UK	(a)	-1.27	*(0.18)*	0.15	*(0.04)*				
	(b)	-1.08	*(0.21)*	0.08	*(0.04)*	-0.04	*(0.04)*	0.08	*(0.02)*
US	(a)	-0.73	*(0.15)*	0.31	*(0.04)*				
	(b)	-0.37	*(0.17)*	0.35	*(0.04)*	0.12	*(0.04)*	0.04	*(0.02)*

Notes: This table shows the estimated parameters for the extended HP model and their standard deviation (in parentheses). Rows (a) show the estimations of the baseline model, while rows (b) present the estimations of the model augmented with credit and house prices.

Table 4.8: Parameter estimates: semi-structural model (part 1)

	model	α_1		α_2		$\bar{g}^y$		β_1		β_2		γ_1	
Prior density type		beta		beta		normal		beta		gamma		beta	
Prior	(a)	0.6	*(0.2)*	0.1	*(0.05)*	$\overline{\Delta y_t}$*	*(0.001)*	0.6	*(0.2)*	0.3	*(0.2)*	0.6	*(0.2)*
	(b)	0.6	*(0.2)*	0.1	*(0.05)*	$\overline{\Delta y_t}$*	*(0.001)*	0.6	*(0.2)*	0.3	*(0.2)*	0.6	*(0.2)*
Posterior													
CA	(a)	0.9013	*(0.0508)*	0.1271	*(0.0007)*	0.0066	*(0.0434)*	0.9034	*(0.0501)*	0.1107	*(0.0492)*	0.8474	*(0.0757)*
	(b)	0.9067	*(0.0480)*	0.1255	*(0.0470)*	0.0067	*(0.0007)*	0.9054	*(0.0755)*	0.1098	*(0.0433)*	0.8312	*(0.0531)*
FR	(a)	0.9038	*(0.0470)*	0.1303	*(0.0007)*	0.0048	*(0.0523)*	0.9104	*(0.0479)*	0.1046	*(0.0516)*	0.8633	*(0.0732)*
	(b)	0.9024	*(0.0555)*	0.1333	*(0.0558)*	0.0049	*(0.0007)*	0.9076	*(0.0766)*	0.1112	*(0.0502)*	0.8571	*(0.0445)*
DE	(a)	0.8545	*(0.0491)*	0.1280	*(0.0008)*	0.0050	*(0.0600)*	0.9067	*(0.0491)*	0.0883	*(0.0418)*	0.8959	*(0.0604)*
	(b)	0.8676	*(0.0522)*	0.1331	*(0.0415)*	0.0051	*(0.0007)*	0.9021	*(0.0741)*	0.0910	*(0.0422)*	0.8801	*(0.0497)*
IT	(a)	0.8735	*(0.0289)*	0.0693	*(0.0008)*	0.0045	*(0.0567)*	0.9443	*(0.0302)*	0.1782	*(0.0702)*	0.9087	*(0.0472)*
	(b)	0.8880	*(0.0319)*	0.0712	*(0.0596)*	0.0044	*(0.0008)*	0.9385	*(0.0527)*	0.1739	*(0.0335)*	0.8972	*(0.0478)*
JP	(a)	0.8670	*(0.0223)*	0.0443	*(0.0009)*	0.0053	*(0.0625)*	0.8430	*(0.0800)*	0.1173	*(0.0595)*	0.8458	*(0.0866)*
	(b)	0.8854	*(0.0773)*	0.0519	*(0.0534)*	0.0054	*(0.0009)*	0.8450	*(0.0917)*	0.1072	*(0.0405)*	0.8228	*(0.0498)*
ES	(a)	0.9359	*(0.0391)*	0.1020	*(0.0009)*	0.0055	*(0.0342)*	0.8928	*(0.0654)*	0.1023	*(0.0513)*	0.8821	*(0.0511)*
	(b)	0.9464	*(0.0549)*	0.1254	*(0.0340)*	0.0055	*(0.0008)*	0.9044	*(0.0638)*	0.0671	*(0.0625)*	0.8321	*(0.0504)*
UK	(a)	0.8914	*(0.0472)*	0.1292	*(0.0008)*	0.0056	*(0.0567)*	0.8573	*(0.0814)*	0.1145	*(0.0661)*	0.8200	*(0.0903)*
	(b)	0.9223	*(0.0843)*	0.1505	*(0.0504)*	0.0056	*(0.0007)*	0.8422	*(0.0994)*	0.0940	*(0.0560)*	0.7697	*(0.0517)*
US	(a)	0.8888	*(0.0473)*	0.1301	*(0.0007)*	0.0069	*(0.0499)*	0.8988	*(0.0483)*	0.1422	*(0.0575)*	0.8565	*(0.0653)*
	(b)	0.8986	*(0.0440)*	0.1399	*(0.0535)*	0.0069	*(0.0007)*	0.9123	*(0.0714)*	0.1256	*(0.0492)*	0.8478	*(0.0480)*

*Notes: This table shows the estimated parameters for the semi-structural model and their standard deviation (in parentheses). Rows (a) show the estimations of the baseline model, while rows (b) present the estimations of the model augmented with credit and house prices. * $\overline{\Delta y_t}$ is the country-level average GDP growth rate.*

Table 4.9: Parameter estimates: semi-structural model (part 2)

	model	γ_2		γ_3		κ_1		κ_2		ϕ_1^{cr}		ϕ_2^{cr}	
Prior density type		gamma		beta		beta		gamma		beta		normal	
Prior	(a)	0.3	*(0.1)*	0.9	*(0.05)*	0.6	*(0.2)*	0.3	*(0.2)*				
	(b)	0.3	*(0.1)*	0.9	*(0.05)*	0.6	*(0.2)*	0.3	*(0.2)*	0.6	*(0.2)*	0.1	*(0.2)*
Posterior													
CA	(a)	0.1664	*(0.0472)*	0.9035	*(0.0475)*	0.7809	*(0.1177)*	0.3384	*(0.2285)*				
	(b)	0.1614	*(0.0828)*	0.8989	*(0.1529)*	0.8014	*(0.0218)*	0.2644	*(0.0696)*	0.9593	*(0.0549)*	0.2521	*(0.0008)*
FR	(a)	0.1770	*(0.0511)*	0.8993	*(0.0492)*	0.8279	*(0.0691)*	0.2252	*(0.1130)*				
	(b)	0.1802	*(0.0780)*	0.9031	*(0.1385)*	0.7965	*(0.0287)*	0.2787	*(0.0870)*	0.9409	*(0.0508)*	0.2185	*(0.0008)*
DE	(a)	0.1429	*(0.0411)*	0.8992	*(0.0490)*	0.7876	*(0.0906)*	0.4137	*(0.1909)*				
	(b)	0.1436	*(0.0804)*	0.8974	*(0.1747)*	0.7897	*(0.0245)*	0.3703	*(0.0658)*	0.9652	*(0.0340)*	0.1148	*(0.0009)*
IT	(a)	0.1399	*(0.0389)*	0.9023	*(0.0465)*	0.6617	*(0.1094)*	0.4326	*(0.1647)*				
	(b)	0.1311	*(0.1185)*	0.9038	*(0.1600)*	0.6983	*(0.0257)*	0.3106	*(0.0840)*	0.9562	*(0.0305)*	0.1581	*(0.0009)*
JP	(a)	0.1552	*(0.0442)*	0.8989	*(0.0504)*	0.7015	*(0.0905)*	0.8191	*(0.3036)*				
	(b)	0.1447	*(0.0955)*	0.9017	*(0.3354)*	0.6785	*(0.0220)*	0.8654	*(0.0908)*	0.9645	*(0.0368)*	0.0951	*(0.0009)*
ES	(a)	0.2205	*(0.0601)*	0.9002	*(0.0493)*	0.7585	*(0.0960)*	0.1906	*(0.0969)*				
	(b)	0.2248	*(0.0967)*	0.9010	*(0.0818)*	0.7418	*(0.0168)*	0.1531	*(0.0779)*	0.9665	*(0.0407)*	0.3039	*(0.0010)*
UK	(a)	0.1998	*(0.0608)*	0.9031	*(0.0491)*	0.7851	*(0.0972)*	0.1932	*(0.1102)*				
	(b)	0.1853	*(0.0837)*	0.9000	*(0.0701)*	0.8007	*(0.0250)*	0.1318	*(0.0923)*	0.9498	*(0.0484)*	0.3686	*(0.0010)*
US	(a)	0.1837	*(0.0470)*	0.9058	*(0.0460)*	0.7142	*(0.0868)*	0.5280	*(0.1687)*				
	(b)	0.1833	*(0.0889)*	0.8982	*(0.1692)*	0.7305	*(0.0269)*	0.4823	*(0.0679)*	0.9540	*(0.0368)*	0.2399	*(0.0009)*

Notes: see notes from Table 4.8.

Table 4.10: Parameter estimates: semi-structural model (part 3)

	model	ϕ_3^{cr}		$\bar{g}^{cr}$		ϕ_1^{hp}		ϕ_2^{hp}		ϕ_3^{hp}		$\bar{g}^{hp}$	
Prior density type		beta		normal		beta		normal		beta		normal	
Prior	(b)	0.1	*(0.05)*	$\overline{\Delta cr_t}$*	*(0.001)*	0.6	*(0.2)*	0.3	*(0.2)*	0.1	*(0.05)*	$\overline{\Delta hp_t}$**	*(0.001)*
Posterior													
CA	(b)	0.1309	*(0.0331)*	0.0116	*(0.0435)*	0.9324	*(0.0017)*	0.2805	*(0.0013)*	0.1171	*(0.0014)*	0.0067	*(0.0421)*
FR	(b)	0.1307	*(0.0223)*	0.0084	*(0.0510)*	0.9532	*(0.0013)*	0.3702	*(0.0007)*	0.1008	*(0.0009)*	0.0048	*(0.0443)*
DE	(b)	0.0778	*(0.0204)*	0.0064	*(0.0456)*	0.9624	*(0.0011)*	0.1631	*(0.0007)*	0.1235	*(0.0005)*	-0.0001	*(0.0566)*
IT	(b)	0.0821	*(0.0180)*	0.0080	*(0.0315)*	0.9679	*(0.0013)*	0.4705	*(0.0009)*	0.1193	*(0.0011)*	0.0037	*(0.0431)*
JP	(b)	0.0631	*(0.0327)*	0.0046	*(0.0305)*	0.9530	*(0.0013)*	0.1536	*(0.0008)*	0.0801	*(0.0008)*	-0.0002	*(0.0630)*
ES	(b)	0.0818	*(0.0305)*	0.0087	*(0.0461)*	0.9320	*(0.0018)*	0.4258	*(0.0011)*	0.1004	*(0.0017)*	0.0052	*(0.0274)*
UK	(b)	0.1131	*(0.0343)*	0.0119	*(0.0576)*	0.9089	*(0.0023)*	0.4975	*(0.0015)*	0.1099	*(0.0016)*	0.0093	*(0.0406)*
US	(b)	0.0993	*(0.0278)*	0.0084	*(0.0450)*	0.9489	*(0.0012)*	0.1869	*(0.0008)*	0.0911	*(0.0009)*	0.0040	*(0.0422)*

*Notes: see notes from Table 4.8. * $\overline{\Delta cr_t}$ is the country-level average credit growth rate.** $\overline{\Delta hp_t}$ is the country-level average house prices growth rate.*

Table 4.11: Parameter estimates: semi-structural model (part 4)

	model	$\varepsilon_t^{c^*}$		$\varepsilon_t^{\hat{c}}$		$\varepsilon_t^{\pi^*}$		$\varepsilon_t^{\hat{\pi}}$		$\varepsilon_t^{u^*}$		$\varepsilon_t^{\hat{u}}$	
Prior density type		invgamma		invgamma		invgamma		invgamma		invgamma		invgamma	
Prior	(a)	0.1	*(∞)*	2	*(∞)*	0.1	*(∞)*	2	*(∞)*	0.1	*(∞)*	2	*(∞)*
	(b)	0.1	*(∞)*	2	*(∞)*	0.1	*(∞)*	2	*(∞)*	0.1	*(∞)*	2	*(∞)*
Posterior													
CA	(a)	0.0074	*(0.0010)*	0.0204	*(0.0007)*	0.0053	*(0.0005)*	0.0146	*(0.0008)*	0.0047	*(0.0010)*	0.0130	*(0.0006)*
	(b)	0.0074	*(0.1093)*	0.0208	*(0.0493)*	0.0051	*(0.0006)*	0.0146	*(0.0010)*	0.0046	*(0.0005)*	0.0131	*(0.0012)*
FR	(a)	0.0065	*(0.0011)*	0.0186	*(0.0006)*	0.0055	*(0.0004)*	0.0151	*(0.0008)*	0.0051	*(0.0012)*	0.0145	*(0.0007)*
	(b)	0.0064	*(0.1210)*	0.0184	*(0.0403)*	0.0055	*(0.0008)*	0.0152	*(0.0011)*	0.0052	*(0.0006)*	0.0143	*(0.0010)*
DE	(a)	0.0070	*(0.0011)*	0.0194	*(0.0007)*	0.0048	*(0.0005)*	0.0134	*(0.0006)*	0.0046	*(0.0010)*	0.0129	*(0.0006)*
	(b)	0.0071	*(0.0759)*	0.0195	*(0.0532)*	0.0048	*(0.0007)*	0.0134	*(0.0010)*	0.0046	*(0.0005)*	0.0129	*(0.0012)*
IT	(a)	0.0071	*(0.0010)*	0.0196	*(0.0007)*	0.0058	*(0.0004)*	0.0160	*(0.0009)*	0.0045	*(0.0011)*	0.0127	*(0.0006)*
	(b)	0.0072	*(0.0920)*	0.0201	*(0.0452)*	0.0058	*(0.0006)*	0.0159	*(0.0009)*	0.0045	*(0.0005)*	0.0126	*(0.0011)*
JP	(a)	0.0124	*(0.0013)*	0.0354	*(0.0009)*	0.0058	*(0.0003)*	0.0163	*(0.0010)*	0.0054	*(0.0013)*	0.0152	*(0.0008)*
	(b)	0.0125	*(0.0880)*	0.0348	*(0.0369)*	0.0058	*(0.0009)*	0.0164	*(0.0012)*	0.0053	*(0.0006)*	0.0150	*(0.0026)*
ES	(a)	0.0074	*(0.0012)*	0.0217	*(0.0007)*	0.0065	*(0.0005)*	0.0179	*(0.0010)*	0.0057	*(0.0013)*	0.0157	*(0.0009)*
	(b)	0.0075	*(0.1457)*	0.0216	*(0.0384)*	0.0065	*(0.0008)*	0.0179	*(0.0012)*	0.0057	*(0.0007)*	0.0156	*(0.0014)*
UK	(a)	0.0081	*(0.0016)*	0.0242	*(0.0009)*	0.0066	*(0.0005)*	0.0190	*(0.0011)*	0.0065	*(0.0017)*	0.0184	*(0.0011)*
	(b)	0.0083	*(0.1314)*	0.0243	*(0.0439)*	0.0066	*(0.0011)*	0.0191	*(0.0017)*	0.0064	*(0.0009)*	0.0184	*(0.0015)*
US	(a)	0.0060	*(0.0009)*	0.0165	*(0.0006)*	0.0053	*(0.0004)*	0.0149	*(0.0007)*	0.0046	*(0.0010)*	0.0129	*(0.0007)*
	(b)	0.0062	*(0.1105)*	0.0166	*(0.0294)*	0.0053	*(0.0006)*	0.0151	*(0.0010)*	0.0046	*(0.0005)*	0.0129	*(0.0010)*

Notes: see notes from Table 4.8.

Table 4.12: Parameter estimates: semi-structural model (part 5)

	model	ε_t^{gu}		$\varepsilon_t^{y^*}$		$\varepsilon_t^{\hat{y}}$		ε_t^{gy}	
Prior density type		invgamma		invgamma		invgamma		invgamma	
Prior	(a)	0.01	*(∞)*	0.01	*(∞)*	2	*(∞)*	0.01*	*(∞)*
	(b)	0.01	*(∞)*	0.01	*(∞)*	2	*(∞)*	0.01*	*(∞)*
Posterior									
CA	(a)	0.0019	*(0.0010)*	0.0022	*(0.0005)*	0.0147	*(0.0013)*	0.0018	*(0.0015)*
	(b)	0.0019	*(0.0015)*	0.0022	*(0.0011)*	0.0147	*(0.0007)*	0.0017	*(0.0005)*
FR	(a)	0.0021	*(0.0011)*	0.0020	*(0.0006)*	0.0140	*(0.0010)*	0.0015	*(0.0013)*
	(b)	0.0021	*(0.0014)*	0.0020	*(0.0012)*	0.0139	*(0.0008)*	0.0015	*(0.0005)*
DE	(a)	0.0019	*(0.0009)*	0.0024	*(0.0005)*	0.0161	*(0.0012)*	0.0019	*(0.0016)*
	(b)	0.0019	*(0.0015)*	0.0023	*(0.0010)*	0.0162	*(0.0007)*	0.0019	*(0.0004)*
IT	(a)	0.0019	*(0.0009)*	0.0022	*(0.0005)*	0.0153	*(0.0011)*	0.0016	*(0.0014)*
	(b)	0.0019	*(0.0015)*	0.0022	*(0.0012)*	0.0153	*(0.0008)*	0.0015	*(0.0007)*
JP	(a)	0.0022	*(0.0012)*	0.0026	*(0.0007)*	0.0180	*(0.0027)*	0.0010	*(0.0028)*
	(b)	0.0022	*(0.0029)*	0.0026	*(0.0013)*	0.0182	*(0.0009)*	0.0011	*(0.0007)*
ES	(a)	0.0022	*(0.0012)*	0.0023	*(0.0006)*	0.0156	*(0.0013)*	0.0021	*(0.0016)*
	(b)	0.0023	*(0.0015)*	0.0023	*(0.0013)*	0.0158	*(0.0008)*	0.0021	*(0.0008)*
UK	(a)	0.0026	*(0.0016)*	0.0025	*(0.0009)*	0.0180	*(0.0015)*	0.0019	*(0.0021)*
	(b)	0.0027	*(0.0020)*	0.0024	*(0.0019)*	0.0182	*(0.0011)*	0.0016	*(0.0009)*
US	(a)	0.0018	*(0.0010)*	0.0020	*(0.0005)*	0.0147	*(0.0009)*	0.0017	*(0.0013)*
	(b)	0.0019	*(0.0013)*	0.0021	*(0.0010)*	0.0148	*(0.0006)*	0.0018	*(0.0005)*

*Notes: see notes from Table 4.8. * 0.001 for JP*

Table 4.13: Parameter estimates: semi-structural model (part 6)

	model	$\varepsilon_t^{cr^*}$		$\varepsilon_t^{\hat{cr}}$		ε_t^{gcr}		$\varepsilon_t^{hp^*}$		$\varepsilon_t^{\hat{hp}}$		ε_t^{ghp}	
Prior density type		invgamma		invgamma		invgamma		invgamma		invgamma		invgamma	
Prior	(b)	0.01	*(∞)*	2	*(∞)*	0.01	*(∞)*	2	*(∞)*	0.01	*(∞)*	2	*(∞)*
Posterior													
CA	(b)	0.0022	*(0.0004)*	0.0157	*(0.0007)*	0.0018	*(0.0010)*	0.0035	*(0.0009)*	0.0279	*(0.0010)*	0.0036	*(0.0007)*
FR	(b)	0.0024	*(0.0004)*	0.0159	*(0.0009)*	0.0017	*(0.0011)*	0.0024	*(0.0010)*	0.0171	*(0.0010)*	0.0031	*(0.0006)*
DE	(b)	0.0022	*(0.0005)*	0.0146	*(0.0007)*	0.0018	*(0.0009)*	0.0022	*(0.0008)*	0.0153	*(0.0011)*	0.0019	*(0.0008)*
IT	(b)	0.0026	*(0.0004)*	0.0179	*(0.0009)*	0.0029	*(0.0011)*	0.0027	*(0.0009)*	0.0195	*(0.0010)*	0.0035	*(0.0007)*
JP	(b)	0.0026	*(0.0004)*	0.0170	*(0.0009)*	0.0025	*(0.0013)*	0.0024	*(0.0009)*	0.0176	*(0.0013)*	0.0026	*(0.0009)*
ES	(b)	0.0026	*(0.0005)*	0.0181	*(0.0010)*	0.0028	*(0.0014)*	0.0030	*(0.0010)*	0.0238	*(0.0012)*	0.0061	*(0.0007)*
UK	(b)	0.0030	*(0.0004)*	0.0232	*(0.0011)*	0.0025	*(0.0017)*	0.0036	*(0.0009)*	0.0274	*(0.0016)*	0.0038	*(0.0008)*
US	(b)	0.0021	*(0.0004)*	0.0143	*(0.0008)*	0.0022	*(0.0010)*	0.0024	*(0.0010)*	0.0169	*(0.0010)*	0.0048	*(0.0006)*

Notes: see notes from Table 4.8.

4.D List of recessions

Recession is defined as two consecutive quarters of negative seasonally adjusted GDP growth. "Mild recessions" are those which last less than a year *and* result in a drop in GDP of less than 1 per cent. The list of recessions in our sample is (mild recessions are in parentheses):

- CA: (1980q2–1980q3), 1981q3–1982q4, 1990q2–1991q1, 2008q4–2009q2, (2015q1–2015q2)
- FR: 1992q2–1993q1, 2008q2–2009q2
- DE: 1974q4–1975q2, 1980q2–1980q4, 1982q2–1982q3, (1991q2–1992q3), 1992q2–1993q1, 1995q4–1996q1, 2002q4–2003q1, 2008q2–2009q1, (2012q4–2013q1)
- IT: 1974q4–1975q2, 1977q2–1977q3, 1982q2–1982q4, 1992q2–1993q3, (2001q2–2003q2), (2007q3–200 2008q2–2009q2, 2011q3–2013q1
- JP: 1993q2–1993q3, 1998q1–1998q2, 2001q2–2001q4, (2007q2–2007q3), 2008q2–2009q1, 2010q4–2011 (2012q2–2012q3), 2014q2–2014q3
- ES: (1978q3–1979q1), (1981q1–1981q2), 1992q4–1993q2, 2008q3–2009q4, 2011q1–2013q3
- UK: 1990q3–1991q3, 2008q2–2009q2
- US: 1974q3–1975q1, 1980q2–1980q3, 1981q4–1982q1, 2008q3–2009q2

The list of recessions is consistent with general knowledge about economic and financial turmoil in these countries in our sample. For the US, the main recessions identified by the NBER[20] are listed: the oil crisis of 1974; a relatively short recession in 1980; the recession caused by the sharp increase in oil prices following the Iranian revolution and the tight monetary policy in the US in 1981; and the Great Financial Crisis of 2008-2009. The NBER's recession following the collapse of the speculative dot-com bubble in 2001 is excluded from our analysis because it did not last 2 consecutive quarters. As for the Euro Area countries, the financial recessions identified by Lang et al. (2019) are included in our list. Other recessions which do not result from systemic financial crises are: the recessions in the 1980s and the 1990s in DE; the recessions in the 1970s and the 1980s in IT; and the recession in the 1990s in ES. All recessions listed in Lang et al. (2019) for CA, JP and the UK are also included in our list.

[20] See https://www.nber.org/cycles/.

4.E Distance between the time of entry into recession and the closest peak

A peak in the output gap is identified if $\hat{y}_{it} > \hat{y}_{it+l}, \forall l \in \{-12, ..., -1, 1, ..., 12\}$. The histogram of the distances between the time of entry into recession and the closest peak is plotted in Figure 4.6. We disregard the observations in which the entry into recession precedes the closest peak by more than four quarters (15 cases out of the total of 204 distance measures).

Figure 4.6: Distance between the time of entry into recession and the closest peak (all countries and models)