What drives pension reforms in the OECD?[1]

Roel Beetsma[2], Franc Klaassen,[3] Ward Romp[4] and Ron van Maurik[5]

Abstract

Based on narrative identification we construct a novel comprehensive dataset of pension reform measures in OECD countries from 1970 until 2017. We then study the timing of these measures. Our main and new result is that business cycle indicators are important for their timing: a worsening makes contractionary measures more likely and expansionary measures less likely. The demography matters only in the sense that the OECD-wide demography explains the general reform trend for a country. We find no evidence that country-specific or short-run demographic developments matter. We discuss a conceptual framework with adjustment costs of changing pension generosity that can account for both the reform responsiveness to the business cycle and the lack of responsiveness to changes in demographic forecasts. We also discuss potential policy implications of our findings.

Final version: January 2020

JEL codes: H55, H62, J11, J26.

Keywords: pension reform measures, narrative identification, expansion, contraction, old-age dependency ratio, business cycle indicators.

---

[1] A first version of this paper based on a subset of the current dataset was circulated as CEPR Discussion Paper 12313, dated 19 September 2017. We thank the editors, three anonymous referees, a discussant, Ole Beier Sorensen, Robert Holzmann, participants at the 70th Economic Policy Panel Meeting (Helsinki, 10-11 October 2019) and seminar participants at the National Bank of Poland, the Dutch Central Bank, the University of York, the Fisk Workshop on Fiscal Policy and Ageing organized by the Austrian Fiscal Advisory Council (October 6, 2017), the 12th Annual Meeting of the Portuguese Economic Journal (Lisbon, July 6, 2018), the OECD, the PeRCent Annual Conference 2018 (Copenhagen, May 30, 2018) and the 7th workshop on empirical macroeconomics (Ghent, June 6-7, 2019). This research was made possible with the financial support of pension and wealth manager MN.

[2] Corresponding author. Address: MN Chair in Pension Economics, Amsterdam School of Economics, University of Amsterdam, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands; phone: +31.20.5255280; e-mail: r.m.w.j.beetsma@uva.nl; European Fiscal Board; CEPR; CESIfo; Netspar; Tinbergen Institute.

[3] Address: Amsterdam School of Economics, University of Amsterdam.

[4] Address: Amsterdam School of Economics, University of Amsterdam.

[5] Address: Amsterdam School of Economics, University of Amsterdam.

1 Introduction

Reform measures to enhance the financial sustainability of pension arrangements are high on the agendas of national policymakers as well as of international organizations. So far, however, systematic empirical investigation of what determines the *timing* of pension reform measures is in short supply. This is unfortunate, because the outcomes of such an analysis may provide insights into the circumstances that are most conducive to the successful implementation of pension reforms.

This paper makes two contributions to the literature on pension reform. First, following a "narrative approach" based on various international databases, OECD publications and national sources, we construct a new dataset on pension reform measures in OECD countries over the period 1970 – 2017. We document these reforms and categorize them, indicating whether they expand or contract pension arrangements, and whether broad or narrow groups are affected. To the best of our knowledge this is the most comprehensive dataset on pension reform measures that currently exists, covering a large set of countries over a long sample period. We select pension reform measures that we expect to affect both the generosity of a pension arrangement and the intertemporal government budget constraint. The dataset is "real-time" in the sense that we assign each measure to the year in which it is legislated. Further, the dataset differs from existing datasets in that it focuses on the countries that have been OECD member (almost) from the start. So far, research has mostly focused on pension reforms in Latin America and the Central and Eastern European countries (e.g., see Madrid, 2002; Brooks, 2007b; and Orenstein, 2005, 2013). These reforms are dominated by the privatizations in the 1980s and 1990s, often following the early example of Chile.

Our second, and main, contribution is that we show that projected OECD-wide demographic trends can explain general reform trends in a country, but not the *timing* of reform measures. Business cycle indicators – broadly defined, so capturing economic growth, unemployment and the state of the public finances – *do* have a significant effect on the timing of reform measures: a worsening of the business cycle makes contractionary measures more likely and expansionary measures less likely. This is a result that, to the best of our knowledge, has not been established empirically before. Our interpretation of the evidence is that projections of population ageing in the OECD indicate the pressure to undertake reform measures, while the business cycle determines their timing. Compared to demographic projections, the business cycle fluctuates at a much higher frequency, thereby creating the conditions in which it is optimal to adopt reform measures, before mounting pressures from demographic projections have a chance to do so. Anecdotal evidence supports our findings. After a decades-long stagnation of the debate in the Netherlands, under pressure of the economic and financial crisis only a few weeks were needed in 2012 to decide on a schedule to gradually increase the public-pension retirement age. In fact, recently, in 2019, with the economy at full capacity and the government running a surplus, parliament adopted a law to slow down the pace at which the retirement age for the public pension is increasing. In 2011, France and the UK already decided to raise the retirement age more rapidly than originally planned, while in 2013 Spain decided to raise the retirement age. The period 2011-2013 is the period of Eurozone-wide fiscal consolidation in the wake of the eruption of the zone's debt crisis. We also investigate the role of political variables and economic and financial crises, but these variables have at most limited explanatory power, while we are unable to detect systematic relationships between these variables and the reform regimes. For the political variables this is in line with Duval et al.'s (2018) findings for labor and product market reforms in advanced economies. Most importantly, the main findings alluded to above continue to hold.

We discuss a simple conceptual framework featuring fixed costs of pension reform that is capable of replicating the result that the timing of pension reform measures depends on the business cycle, but not on changes in the (projected) old-age dependency ratio. Such a framework differs fundamentally from existing frameworks in the related literature, which focus mostly on the size of the pension system or its constituent pillars.

While there is little to no literature on the precise timing of pension reforms, our paper connects to various other strands in the pension literature. First, while systematic econometric analysis of pension reform measures is rare, two other recent papers have compiled datasets of pension reforms. Verbic and Spruk (2019) construct a dataset for 34 countries over the period 1970-2013, documenting the pension pillar that is affected by the reform, but not the budgetary effect of the reform. Leibrecht and Fong (2017) document the year in which a country introduced a second-pillar defined-contribution scheme. Verbic and Spruk (2019) use their dataset to relate political indicators to the transition from unfunded to funded pensions, while Leibrecht and Fong focus on the political, economic and social determinants of retirement income privatization. In contrast, our focus is on the timing of pension reform measures that affect both public and individual resources (positively or negatively), of which the state of the business cycle turns out to be the most important determinant, while political variables appear to be of at most limited relevance.

Our paper relates also to other work exploring the determinants of (changes in) pension arrangements. The literature has suggested a number of plausible driving factors of such changes.[6] First, there is the potential role of demography. Persson and Tabellini (2000) describe two opposing effects of a higher old-age dependency ratio on the size of a PAYG system. In an older society, on the

[6] Other suggested factors than those discussed are the availability of information and social dialogue (e.g., Boeri and Tabellini, 2012), peer adoption (e.g., Brooks 2007a) and political factors (e.g., Giuliano et al., 2013).

one hand the rate of return on contributions to a PAYG system is lower, making the system less attractive, while on the other hand population ageing enhances the political weight of the elderly, making it harder for politicians to engage in contractionary reform. Other studies (e.g. Gonzales-Eiras and Niepelt, 2008) relate changes in social security to its intergenerational risk-sharing aspects (see also D'Amato and Galasso, 2010). Empirically, however, the role of the demography is not so clear-cut (e.g., Blinder and Krueger, 2004). While theory suggests that demography is an important determinant for PAYG pension reform, the empirical evidence is weak. In fact, for the U.S. and Western Europe, Razin *et al*. (2002) even find a negative correlation between the old-age dependency ratio and the generosity of social security transfers. Second, the size of the implicit pension debt is a potential determinant of reform of PAYG defined-benefit pensions (James and Brooks, 2001). Third, external constraints, such as those imposed by Europe's Stability and Growth Pact, may stimulate pension reform. Bertola and Boeri (2002) argue that such constraints may have stimulated a reduction in the generosity of social security after 1997. Fourth, pension arrangements can be highly distortionary, leading employees to work less or retire earlier than under a system that gives stronger incentives for work – see, e.g., the contributions in Gruber and Wise (2009). The correction of such distortions may be another reason for reform. Finally, there is the role of ideology. Pension privatization in Latin America was stimulated by the paradigm shift towards neo-liberalization inspired by Thatcherism and the promotion of private pensions by international organizations such as the World Bank (see World Bank, 1994, Brooks, 2007b, and Orenstein, 2005 and 2013), which also emphasized the benefits of capital market deepening, increased private savings and higher economic growth.

Our main empirical finding – the timing of pension reform measures is linked to the business cycle – is related to the crisis-induced reform literature (Rodrik, 1996, Abiad and Mody, 2005, Bonfiglioli and Gancia, 2015, Ranciere and Tornell, 2015, Mahmalat and Curran, 2018),[7] but differs in two fundamental ways. This literature finds that structural reforms are typically legislated during periods of poor economic performance, but it tends to focus on financial liberalization, trade liberalization, and inflation and sovereign indebtedness issues. We focus on pension reform measures. These reform measures are special, since sustainability issues associated with pension provision due to future demographic changes are known well in advance. The "regular" crisis-induced reform literature focuses on contemporaneous rather than anticipated future crises. This may not be surprising: this literature suggests that an economic crisis may be a particularly suitable moment for reform, because only then policymakers become sufficiently aware of the need to fix

[7] Campos et al. (2010) find a relatively large role for political crises in determining labor market and trade liberalization reforms.

structural deficiencies through fundamental reform (Tommasi and Velasco, 1996, and Tommasi, 2017). A second difference is that our dataset allows us to focus on both contractionary and expansionary reform measures. We find that the timing of contractionary reform measures is related to a cyclically-weak economy. Expansionary reform measures, which include such structural reform measures as increased coverage of women, tend to be implemented during economic upswings. Such expansionary reform measures are not considered in the crisis-induced literature.[8]

The remainder of the paper is organized as follow. Section 2 presents the data and summarizes the construction of our dataset of pension reform measures. Section 3 sets up the empirical framework, while Section 4 reports and discusses the estimation results. Section 5 discusses the conceptual framework that can rationalize our empirical findings. Section 6 concludes the main text and discusses potential policy implications of our findings. Robustness and additional empirical results are reported in the online Appendix.

2 The data

2.1 Pension reform measures and pension reform regimes

Our database spans the 23 initial OECD countries over the period 1970 until 2017 and covers all reform measures, both smaller (parametric) and more fundamental, to pension arrangements that affect both the present value of retirement income and the government's intertemporal budget constraint.[9] The identification of reform measures is based on a narrative approach. In each year, and for each country, we list the changes in pension arrangements based on a careful reading of records or documents from four main databases: the NATLEX database of the International Labor Organization (ILO, 2019), the International Social Security Association (ISSA, 2019) database, the OECD, and the European Commission's LABREF (2019). We supplement these with data from ad-hoc sources. A detailed description of the data collection and the allocation of observations into categories is provided in Appendix A.[10]

We are interested in measures that we expect to eventually affect the intertemporal government budget constraint, because these are of particular importance for legislators. Changes that take place purely in the private sector sphere are not included. We date reform measures

[8] There is a literature on political and legal constraints that prevent policymakers from pursuing reform in normal times, but that become softer during a crisis. See, for example, Swagel's (2015) recount of the US Treasury's failure to persuade banks to strengthen their capital position prior to the recent global financial crisis, while bank regulators were able to force banks into recapitalizing themselves after the start of the crisis.

[9] Hence, also changes in retirement age are included (to the extent that this also affects the government's intertemporal resources), because they change the number of years in retirement and, hence, the present value of retirement income.

[10] A comparison with the datasets of Leibrecht and Fong (2017) and Verbic and Spruk (2019) is found in Appendix B.

according to the year in which they are legislated. The reason is that we want to explain reform decisions on the basis of the information that is available at the moment the decision is made. It is obviously conceivable that in many instances the discussion about a reform starts in a year before the reform is legislated, or possibly even earlier. However, it is practically unfeasible to uncover for each reform measure the moment when such a discussion was started. Given that this is not feasible, recording reform measures in deviation from the year of legislation would introduce a source of arbitrariness. It is also possible that implementation of measures starts later than their legislation or that implementation takes place in steps. An example is an increase in the retirement age that is gradually implemented.

Based on our readings of the records we classify reform measures as “Unclear”, when there is no clear description or the description itself is not clear, and “Clear”. The “Clear” ones can be “Not relevant” (for our research) and “Relevant”. The latter we classify along two possible dimensions.

The first dimension concerns the effect on the government’s intertemporal budget constraint, which we categorize as: (1a) “Expanding – Coverage”, reform measures that expand the coverage of the pension arrangement, for example by loosening the eligibility criteria; (1b) “Expanding – Generosity and adequacy”, reform measures that expand the generosity of the pension system, for example by raising the benefit level; (1c) “Expanding – Other”, other measures that have an expanding nature; (2a) “Contracting – Fiscal sustainability”, reform measures that enhance the fiscal sustainability of the pension arrangement, for example by reducing benefits or by raising the retirement age; (2b) “Contracting – Work incentives”, reform measures that enhance work incentives, for example by introducing bonuses for working after the minimum age at which pension benefits can be collected; (2c) “Contracting – Other”, other reform measures that have a contracting nature, and (3) measures for which the budgetary effects are unclear.

The categories “Coverage” and “Generosity and adequacy” can be considered as expansionary of the pension system, because, if a reform measure in either of these two categories is enacted, (long-run) pension obligations will increase. By contrast, the categories “Fiscal sustainability” and “Work incentives” can be considered contractionary, implying a reduction in the (long-run) pension obligations. The majority of the reform measures concern the first, public pillar, while a minority concern the second pillar, but only to the extent that they are expected to influence the intertemporal government finances.

The overlap between our various data sources, and the nature of especially the NATLEX database make it impossible to reliably use the number of reform measures in our database. We can, however, count the number of (country, year) - combinations in which we observe a specific type of reform measure. Table 1 reports the number of (country, year) - combinations with a reform measure per category over the full sample period and for each of the two sub-sample periods when we split the full sample into sub-samples of equal length (1970 – 1993 and 1994 – 2017). In total, we observe 140 (country, year) - combinations with a reform measure that we categorize as "Expanding – Coverage", 247 that we categorize as "Expanding – Generosity and adequacy" and 10 with an "Expanding – Other" reform measure. There are 181 (country, year) - combinations with a "Contracting – Fiscal sustainability" reform measure, 97 with a "Contracting – Work incentives" reform measure and 5 with a "Contracting – Other" reform measure. There is a reasonable balance of expanding measures over the two sub-periods, although the second sub-period features some more of them. By contrast, contracting measures are far more prevalent in the second sub-sample than in the first sub-sample, potentially the result of anticipated demographic pressures on public budgets.

Systematic differences in the frequency of reform measures across the sample countries, likely the result of differences in political culture and institutions, motivate us to explain "reform *regimes*" rather than the likelihood of individual reform *measures*. Therefore, we first create two dummy variables. The dummy "Expansion" is one if at least one reform measure falls within the categories "Expanding – Coverage", "Expanding – Generosity and adequacy" or "Expanding – Other", and zero otherwise. The variable "Contraction" is one if at least one measure falls within the categories "Contracting – Fiscal sustainability", "Contracting – Work incentives", or "Contracting – Other", and zero otherwise. We combine the subcategories into the dummies "Expansion" and "Contraction", because the allocation into these subcategories is not always very sharp. Using these dummies, we define three different reform *regimes*. The first is "Expanding only", which is captured by a dummy equal to one if the dummy "Expansion" is one *and* the dummy "Contraction" is zero. It is zero otherwise. The second regime is "Contracting only", which is captured by a dummy equal to one if the dummy "Expansion" is zero *and* the dummy "Contraction" is one. It is zero otherwise. Finally, there is the regime of "Contracting and expanding", which is captured by a dummy equal to one when *both* dummies "Expansion" and "Contraction" are one, while it is zero otherwise.

The idea of this three-way dissection, and in particular of the definition of a regime "Contracting and expanding", is that governments may buy off public or political resistance to contractionary measures by at the same time expanding the system somewhat in other dimensions (e.g., see Castanheira *et al.*, 2006). This interpretation is supported by the fact that in the majority of the cases when contracting and expanding measures occur in a country in the same year, these reforms are described as a combination of measures in a single text piece in the ISSA documentation. However, the descriptions of the "Contracting and expanding" regimes make clear that in many instances it is the contractionary part that dominates the expansionary part. Moreover, the readings of the various cases suggest that the envisaged degree of contraction regularly exceeds that of the "Contracting only" regimes, consistent with the notion that substantial contractions can be made more acceptable politically by having some compensation elsewhere. Hence, the contracting and expanding components are not expected to offset each other and “Contracting and expanding” is really a regime of its own and is, therefore, indicated by a separate regime dummy.

Figure 1 exhibits the frequencies of the three policy reform regimes in each sample period. In line with Table 1, the figure suggests a mild reduction over time of the frequency of the “Expanding only” regime over the second half of the sample and an upward trend in the frequency of the “Contracting only” and “Contracting and expanding” regimes during the second half of the sample, although from year to year the frequencies fluctuate quite substantially, and in some years the frequencies are zero. Figure 2 depicts the frequency of the different reform regimes for each country. The occurrence of a specific reform regime differs quite substantially across the countries. This may be the result of differences in (political) costs of taking reform measures.

Table 1 also summarizes the information on the frequencies of each reform *regime*. In total, we identify 339 (country, year)-combinations with an expanding reform measure. This is less than 397 - the sum of the subcategories of “Expanding”, implying that for some countries in some years there is more than one expanding measure, while the number of (country, year) - combinations with expanding reform measures *only* is 223. In other words, there are 339-223=116 (country, year) - combinations that fall into the “Contracting and expanding” regime. Further, there are 235 (country, year) – combinations with contracting measures, while the number of (country, year) - combinations with contracting measures *only* is 119.

Table 2 provides further details on the “Contracting and expanding” regime, where we report the number of (country, year) – pairs of the different combinations of expansionary and contractionary measures. The total number of 165 combinations exceeds the number of 116 (country, year) – pairs in which the “Contracting and expanding” regime is observed. The reason is that there are “Contracting and expanding” regimes that contain more than one of the combinations in Table 2.

Table 3 reports the unconditional frequencies over different reform regimes over all observations, as well as the frequencies of each reform regime conditional on a given reform regime having occurred in the previous period. We observe that the unconditional frequencies of specific reform regimes occurring are lower, but not much, than the frequencies conditional on that specific regime having occurred in the previous period, suggesting that the degree of serial correlation in the regimes is minor or that serial correlation absent.

Reform measures can be institutional, such as the replacement of the first by a second pension pillar, or they can be incremental. Naturally, the majority of the reform measures would fall into the latter category. Ideally, we would like to be able to explicitly distinguish between institutional and incremental reform measures. However, the descriptions in our databases usually do not provide enough guidance in this respect. Also, the length of the descriptions appears to provide little guidance regarding the depth of the reform. Yet, from the descriptions it is often clear whether the measures affect specific groups (e.g., farmers only) or broad groups (e.g. all women). Therefore, as our second classification dimension, whenever possible we categorize reform measures whether they affect broad groups of the population or rather specific groups. The former will be referred to as "Many".[11] All the other reform measures, including those for which width of the group is unknown, are referred to as "Not many". Out of the 1104 country-year pairs, 458 feature at least one reform measure. If at least one of these measures is of the "Many" type, then the reform regime is classified as "Many". Otherwise, the reform regime is "Not many". Table 4 reports the frequencies of "Not many" and "Many" of each reform regime. Of the 458 reform regimes, we classify 333 country-year pairs as "Many", and the remaining 125 as "Not many". Of the country-year combinations in the "Expanding only" regime 40% are classified as "Not many". Most (91%) of the "Contracting and expanding" regimes are classified as "Many".

2.2 The demographic variables

As demographic variables we use the current old-age dependency ratio and its 25-year ahead forecast, as well as transformations of these basic variables. The old-age dependency ratio is measured as the number of people of 65 years and older divided by the number of people in the age group 15-64 years. The demographic projections and current data are taken from the various issues of the World Population Prospects of the United Nations. As these are not published in each year, for the missing years we construct the ratio using the projections from publication years. Figure 3 depicts the cross-country mean forecast as well as bands that contain the country-specific forecasts. The mean exhibits a strong upward trend, reflecting the ageing trend in the industrialized world. The country-specific forecasts underlying the bands sometimes deviate substantially from the mean. They are generally quite irregular, potentially reflecting differences in methodology of projections at the country level and the use of new country-specific information when updates are produced. Because such updates can trigger political debate on reform measures, the observed irregularities should not be smoothed out before entering the data into the regression analysis.

Demographic variables are affected by fertility, mortality and migration. The United Nations projections are based on cohort-specific estimates of fertility and mortality. Age and gender profiles of net migration flows are also used; other differences, such as in fertility and mortality rates of migrants, are ignored. However, net migration is typically small. In our sample only Luxembourg

[11] Note that "Many" also includes small, sometimes even temporary changes that affect broad groups in society.

experienced a large net inflow exceeding more than 1% of the total population over multiple 5-year periods. Large outflows exceeding 1% of the population only occurred in the 1950s (Ireland) and 1960s (Portugal), so before the start of our sample.

2.3 The economic and budgetary variables

Our set of economic and budgetary variables comprises per-capita real GDP, inflation as measured by the GDP deflator and the consumer price index, government debt, government revenues, government expenditures, the unemployment rate, the yield on short-term debt, the yield on long-term debt, exports and imports. These variables are mostly taken from the OECD Economic Outlook, the OECD National Accounts, the Ameco dataset (European Commision, 2014), the IMF World Economic Outlook (2019) and the World Bank (2019).

2.4 The political variables

Our political variables are obtained from the Comparative Political Data Set I (Armingeon *et al.*, 2018a, 2018b). The variables we use relate among others to the composition of the cabinet, the composition of the parliament, the political orientation of the government and the parliament, and elections.

2.5 Other variables

Finally, we use the crisis indicators taken from Laeven and Valencia (2018) and dummy variables to indicate the participation (or not) in the European Union (EU) as of 1992,[12] which is the year when the Maastricht Treaty was signed, or the Eurozone as of the year of entry into the zone.

3 The empirical framework

The literature suggests that demographic (Persson and Tabellini, 2000), macroeconomic and budgetary (Thompson, 2009) and political and crisis variables (Drazen and Grilli, 1993, and Tommasi and Velasco, 1996) may affect the propensity to initiate (pension) reform measures. Our baseline specification will be a logit regression that links the occurrence of a reform regime to demographic variables, GDP growth, the public deficit and unemployment. With these baseline variables included, we can explore the role of demographic projections in promoting pension reform measures as well as the role of the state of the economy as captured through different indicators. We thus consider GDP growth, the public deficit and unemployment all as indicators of the current state of the economy.

[12] Our sample does not include countries that entered the EU later.

We include these variables jointly in our regression, because their correlation is far from perfect.[13] For example, labor hoarding generally causes cyclical movements in unemployment to lag behind cyclical movements in output. Also, in the past high unemployment rates have often encouraged the search for alternative channels to shed employees, such as through early retirement. Indeed, Table 5 shows that the correlation between growth and deficits is -0.22, between growth and unemployment is -0.15, and between deficits and unemployment is 0.44.

Our empirical approach deviates in a potentially important way from that suggested by political-economy models, such as Cooley and Soares (1999), Persson and Tabellini (2000) and Tabellini (2000) – see Galasso and Profeta (2002) for a survey. The empirical predictions of these models are usually based on the *current* demographic balance among the cohorts,[14] while in our baseline empirical specification we include demographic *projections*. Equity considerations could lead to contractionary pension measures in order to spread the cost of future increases in the old-age dependency ratio more evenly across the cohorts, in particular by shifting some of the cost also to the cohorts currently alive.

Finally, the baseline also includes a "Maastricht dummy", taking a value of one for all country-year combinations as of 1992 in which a country is a member of the EU, and zero otherwise.[15] The reason for including it, is that supranational budgetary constraints could motivate the adoption of reform measures (Bertola and Boeri, 2002), especially those that improve the public budget in the shorter run, thereby making it easier to adhere to such constraints. A substantial part of our sample concerns European Union (EU) countries that have tried to meet the criteria for accession to the Eurozone and that have been bound by these criteria since they entered the Eurozone.[16] Also, EU countries that do not take part in the common currency are, in principle, bound by these criteria and are supposed to take appropriate measures if they violate these criteria.

We adopt a standard logistic regression specification and model the probability $p_{it,r}$ of country i being in reform regime r ("Expanding only", "Contracting only" and "Contracting and expanding") in year t as:

$$p_{it,r} = \frac{\exp(z_{it,r})}{1+\exp(z_{it,r})}, \tag{1}$$

[13] In fact, regressions based on including each of the business cycle variables separately into the baseline specification yield very similar estimates for the coefficients on the business cycle variables; compare the estimates in Tables E.1-3 in online Appendix E with those in Table 6 below.

[14] Many of these analyses feature a model in which there is a repeated vote about the generosity of the pension system, hence often it is the location of the current median voter that determines the system's generosity.

[15] Elaborate prior experimentation shows that the aforementioned variables are the (only) variables for which there is a systematic role in explaining our reform regimes. The robustness analysis below confirms this.

[16] These criteria constitute a ceiling of 3% on the public deficit – GDP ratio and a ceiling of 60% on the public debt – GDP ratio (e.g., see Beetsma and Uhlig, 1999, and Beetsma and Debrun, 2007).

where $z_{it,r}$ is a reform-regime specific linear function of the explanatory variables:[17]

$$z_{it,r} = \alpha_{0i,r} + \alpha_r{}' BASEVAR_{it} + \delta_r' ADD_{it}, \tag{2}$$

where $\alpha_{0i,r}$ captures country-fixed effects and α_r and δ_r are coefficient vectors of appropriate dimensions. Further,

$$BASEVAR_{it} = \left(\overline{OAD25_t}, OADDEV25_{it}, \Delta OAD25_{it}, GROWTH_{it}, DEF_{it}, UNEMPL_{it}, MAASTRICHT_{it}\right)' \tag{3}$$

is the vector of baseline explanatory variables and ADD_{it} a vector of potential additional variables. Variable $OAD25_{it}$ (not entered directly into the regression – see below) is the projected 25-years ahead old-age dependency ratio, $\overline{OAD25_t}$ is its cross-country average in year t, $OADDEV25_{it} \equiv OAD25_{it} - \overline{OAD25_t}$ is the country-specific deviation from this average and $\Delta OAD25_{it} \equiv OAD25_{it} - OAD25_{i,t-1}$ is its change between years t-1 and t. Instead of including $OAD25_{it}$ directly as an explanatory variable, we allow for a more general formulation by including $\overline{OAD25_t}$ and $OADDEV25_{it}$ separately. This way we can examine whether OECD-wide ageing is a driver of pension reform or whether country-specific deviations in ageing also matter. Observing general demographic trends, international policy organizations like the OECD, the European Commission and the World bank help to raise awareness of the future costs associated with ageing and put (implicit) pressure on national governments to take measures to address these costs. By including $\Delta OAD25_{it}$ we make the formulation even more general and allow for the possibility that a change in a country's projected old-age dependency ratio makes its government more aware of the need to reform pension arrangements. Further, $GROWTH_{it}$ is the GDP growth rate, DEF_{it} is the government's budget deficit as a share of GDP and $UNEMPL_{it}$ is the unemployment rate. Together, these variables are intended to capture the state of the business cycle. Finally, $MAASTRICHT_{it}$ is the aforementioned "Maastricht dummy". All other baseline explanatory variables will be measured as rates or proportions (so, for example, 0.02 instead of 2 percent). We will always include country-fixed effects.[18] These may capture, among other factors, the consequences of cross-country differences in the (political) costs of taking reform measures, leading to systematic differences in the probability of

[17] For simplicity, we will run a separate logistic estimation for all three regimes, so that formally at each data point the alternative to ending up in regime r is to end up in a regime in which no reform measures are taken. However, as our robustness analysis below shows, applying multinomial logit regressions in which the various reform regimes are imposed to be mutually exclusive leaves our results unaffected.

[18] Including country-fixed effects in a panel logit regression produces biased coefficients (Chamberlain, 1980). Given our long sample of 47 years, this bias is presumably small. Still, in Subsection 4.2.7 we show that estimating the model with conditional logit leads to almost identical coefficient estimates.

a reform regime (recall Table 3). In our baseline regression we do not include time-fixed effects, because the robustness analysis below will show that they are not needed.

One may ask whether pension reform measures could themselves have potential feedback effects on the business cycle variables, thereby biasing their coefficient estimates. Such reform measures tend to be phased in over a rather long period and often they are incremental. Hence, most reform measures are unlikely to have material short-run effects on the budget deficit, which is likely dominated by all the other changes in public spending or revenue collection. Short-run effects of reform measures on GDP likely run via the disposable income only of current retirees and, given that the measures are often only incremental, they are unlikely to have a visible effect on a country's GDP. Indeed, the instrumental variables estimates with a linear probability model in Subsection 4.2.7. suggest that feedback effects are not important. This conclusion is supported by the fact that, as we will see in Subsection 4.3, estimates for a split in "Many" and "Not many" yield quite similar results for the latter, for which feedback effects are even less likely than for the full sample.

4 Results

This section describes and interprets the results of our regression analysis. Preliminary analysis suggests that the effect of GDP growth on the probability of the "Expanding only" regime is not stable over time. Tests show that there is a structural break at the end of the 1970s. If we allow for a different intercept and coefficient for economic growth starting in 1980, we find no reason to allow for other breaks, neither in other years, nor for the coefficients of the other business cycle variables. As explained in more detail in Appendix C, we conclude that there is no reason to allow for breaks in the regressions of the other regimes.

4.1 Baseline estimates

Table 6 reports the baseline regression estimates. For each of the three regimes we have two columns, one reporting the coefficient estimate, the other reporting the estimated average marginal effect over all observations of a change in the variable.

We discuss the estimates for the "Expanding only" regime first. Of the demographic variables, the coefficient of the average projected old-age dependency ratio $\overline{OAD25_t}$ is significantly negative at the 5% level, indicating that an OECD-wide upward increase in the projected old-age dependency ratio reduces the likelihood of the "Expanding only" regime.[19] However, the country-specific

[19] Here, and in the sequel, when we refer to a coefficient as "significant" without indicating the specific significance level, we mean that the coefficient is significantly different from zero at the 10% level or higher. Notice that our tests are two-

deviation from the average projection is insignificant and so is the change $\Delta OAD25_{it}$ for an individual country. The irrelevance of the country-specific deviation in the regression may be explained by the fact that the country-specific deviations tend to fluctuate substantially around the mean in Figure 3, which could make it difficult to establish a systematic link between these deviations and the likelihood of a reform regime. Alternatively, or in addition, it could also indicate that it is OECD-wide developments closely related to the general demographic trend, such as increasing awareness of future ageing costs, that contribute to explaining national reform decisions. In view of the break tests discussed above, for "Expanding only" Table 6 reports the coefficient of economic growth both for the period before 1980 and the period starting with 1980. For the latter period, economic growth has a strong, statistically significant, positive effect on the probability of an "Expanding only" regime, indicating that policymakers take the state of the economy into account when considering whether or not to expand pension arrangements: higher GDP growth leads to more expansionary pension measures. A plausible explanation is that higher growth relaxes the public sector budget constraint leading to political pressure to take expansionary measures. Also, workers' wages tend to grow faster during expansionary periods, which likely creates political pressure for benefit recipients to share in the rise in welfare. The insignificant coefficient on the variable $70s \times GROWTH_{it}$ indicates that during the 1970s economic growth has an insignificant effect on the probability of "Expanding only". The break in 1980 may be explained by the 1970s being an atypical period, with two major oil crises exerting a drag on economic growth, while countries were still in a transition phase building up systems of retirement income provision. The (estimates for the) other two cyclical variables, the public deficit and the unemployment rate, are not significant.

The average marginal effect indicates that a one-percentage point higher GDP growth rate raises the likelihood of the "Expanding only" regime by 2.5 percentage points as of 1980. In our sample, the within standard deviation of economic growth is 2.6 percentage points; so a two standard deviation fluctuation (e.g. the difference between a peak in the business cycle and a recession) changes the probability of an expansionary reform by 12.8 percentage points, more than half of the unconditional probability. A one-percentage point higher average forecast of the old-age dependency ratio 25 years from now reduces the likelihood of this regime by 1.0 percentage point. Over our sample, this average forecast increases by 26 percentage points, so this increase lowers the probability of an expansionary reform by 26 percentage points, more than the unconditional probability. The coefficient of the Maastricht Treaty dummy is positive. This positive dummy is driven by the relatively many "Expanding only" regimes between 1994 and 2004 in countries that signed the

---

sided, hence when referring to a coefficient being significantly larger or smaller than zero, we refer to a 5% significance level in that specific direction away from zero.

Maastricht Treaty when compared to the other countries. After 2004 the Maastricht Treaty signatories corrected this expansionary push with a relatively large number of contracting reforms.

For the "Contracting only" regime, the average projected old-age dependency ratio $\overline{OAD25_t}$ has a highly significantly positive effect, suggesting that a country is more likely to adopt contractionary pension measures in response to an OECD-wide increase in the projected old-age dependency ratio, a finding consistent with the negative effect of $\overline{OAD25_t}$ for the "Expanding only" regime. A one-percentage point increase in the forecast of the OECD-average old-age dependency ratio 25 years from now raises the likelihood of the "Contracting only" regime by 0.4 percentage points, smaller in magnitude than the effect on the likelihood of an "Expanding only" regime, but still quantitatively relevant compared to the unconditional probability of 11% (see Table 3). The 26 percentage points increase of the average forecast over the sample period raises the probability 10.4 percentage points. Again, the other demographic variables play no significant role. Economic growth exerts a highly-significant negative effect on contractionary activity, consistent with the highly-significant positive effect found for the "Expanding only" regime, and with an average marginal effect of 1.2. Of the other cyclical variables, the unemployment rate enters with a highly-significant positive coefficient. A one-percentage point increase in the unemployment rate raises the chance of a "Contracting only" regime by 1.5 percentage point. Unemployment has a within standard deviation of 3.0% in our sample, while changes on the order of 6 percentage points are not uncommon. A two standard deviations fluctuation of unemployment changes the probability of observing a "Contracting only" regime by 9.0 percentage points, almost the full unconditional probability. The estimate for the budget deficit is insignificant. The Maastricht dummy enters with a significantly positive sign – this may not be surprising: the run-up to the EU's Economic and Monetary Union and the membership period later coincide with the period in which the "Contracting only" regime tends to prevail. In addition, being eligible for entry into and being member of the EMU requires healthy public finances, and pursuing contracting measures helps in that respect.

Finally, the "Contracting and expanding" regime becomes more likely with an OECD-wide projected increase in the old-age dependency ratio – the coefficient on $\overline{OAD25_t}$ is highly significantly positive. The average marginal effect is comparable to that of the "Contracting only" regime. The other demographic variables are again insignificant. Of the cyclical variables, it is reassuring that the signs of all three estimates are the same as for the "Contracting only" regime and opposite to those for the "Expanding only" regime. This supports the conjecture that a worsening of the cyclical situation induces the government to adopt more reform packages in which substantive contractionary measures are bought off with expansionary measures in other dimensions. It is also consistent with the fact that the public deficit enters with a significantly positive coefficient. The likelihood of this combination regime increases by 0.8 percentage points when the government's budget deficit increases by one percentage point. Relative to an unconditional probability of 11 percent of this regime occurring (see Table 3) this would mean an increase by 7 percent. Finally, we observe that for this combination regime, the average marginal effect of the Maastricht dummy is similar to "Contracting only".

Overall, our baseline estimates suggest that the OECD-wide demographic trend helps to explain the trends in reform regimes, but that it is the cyclical state of the economy that determines the timing of reform measures being taken.

4.2 Robustness

In this subsection we show that the baseline results are robust for a number of variations and extensions of the baseline.

4.2.1 Adding time fixed effects

The baseline regression omits time fixed effects. However, there could be factors other than the OECD-wide ageing projections systematically affecting reforms into a particular direction in all the sample countries in each given year. Hence, we add to the baseline specification of $z_{it,r}$ time fixed effects. This formulation is fully general in the time dimension. For example, it contains as a special case the formulation in which we expand the baseline with a time trend, a specification that could be resorted to if there was a reason to believe that $\overline{OAD25_t}$ does not capture all the effects of trending variables on reforms. Table 7 shows that the coefficients on the business cycle indicators remain largely unchanged (compare with Table 6). Although for space considerations we do not report the estimated time fixed effects, a Wald test that they are jointly zero fails to reject for each of the regimes. Because we have normalized the time fixed effects such that they capture omitted time-specific determinants after accounting for $\overline{OAD25_t}$, their insignificance also implies there is no indication of such omitted determinants. This supports our conclusion that it is OECD-wide ageing projections that affect reforms.

In summary, these results suggest that the baseline regression model without time fixed effects is econometrically adequate so far.

4.2.2 Multinomial logit

The above regressions ignore the fact that our reform regimes, including a "No reform" regime in which no reform measure is undertaken, are mutually exclusive. This fact can be exploited using a multinomial logit model, using the three reform regimes and the "No reform" regime as a dissection of the possible outcome set.[20] Table 8 reports the estimates. We observe that the estimates are similar to those reported in Table 6. As before, the coefficient and average marginal effect on growth are highly-significantly positive for "Expanding only" and highly-significantly negative for "Contracting only", while unemployment remains highly-significantly positive for the latter regime.

4.2.3 Including lags

In this subsection we extend the baseline specification by including lagged values of the reform regime and lags of the explanatory variables. Table 9 reports the estimates when we include the first lag of each of the regime indicators. The estimates confirm the significance of the baseline variables. In two instances is the lagged regime indicator significant. One is in the case of the "Contracting only" regime, where the occurrence of a "Contracting only" regime in the previous period makes this regime less likely in the current period. A possible explanation is that changes are politically costly so politicians prefer to implement these changes in one package. Also consistent with this explanation is the negative effect of a "Contracting and expanding" regime last year on the probability of another one this year. Adding second lags of the reform regime indicators does not have any effect on the estimates.

Our next variant tries to do justice to the fact that, when the circumstances justify some measure, it may take some time before it is turned into law, and hence this may happen only the year after the intention to take some measure is formed. Table 10 includes as explanatory variables the averages of the current and previous year values of the business cycle variables. The estimates are close to those under the baseline and variables that were significant (insignificant) before remain significant (insignificant). We also estimated the model allowing for different coefficients on the current and lagged business cycle variables. For reasons of space we do not report them, but the results are essentially the same.

4.2.4 Additional economic controls

In this subsection we expand the baseline regressions with economic control variables, always including one additional control at a time. The additional controls that we consider are the openness to trade ($OPENNESS_{it}$), the average of the short-term (3-month) and the long-term (10-year) public debt yield ($INTEREST_{it}$), CPI inflation ($INFLATION_{it}$), general government debt as a share of GDP ($DEBT_{it}$) and a dummy $EUROZONE_{it}$, which is one for each Eurozone member as of the year it

[20] The baseline logistic regression specification is now replaced by a more general one that exploits the fact that the reform regimes are mutually exclusive. Concretely, the likelihood $p_{it,r}$ of a reform regime of type r in country i in period t is $p_{it,r} = \frac{\exp(z_{it,r})}{1+\sum_{h=1}^{R}\exp(z_{it,h})}$, where h counts over the set of possible reform regimes "Expanding only", "Contracting only" and "Contracting and expanding". Hence, $R = 3$. The likelihood of ending up in the "No reform" regime is $1 - \sum_{h=1}^{R} p_{it,h}$. The specification of the $z_{it,r}$ is the same as under the baseline, though imposing across the regimes a common break in the intercept and the effect of GDP growth in 1980.

joined the Eurozone and zero, otherwise. As a final additional control, we consider a dummy for the occurrence of a crisis – the estimates with this dummy included are discussed in the next subsection. The estimates with the additional economic controls are reported in Table D.1 in online Appendix D. For the sake of space, we only present the estimated coefficients of the additional variables and we omit the average marginal effects. The significance or insignificance of the business cycle variables is in all instances unaffected. Also numerically, the coefficients on the business cycle variables are always close to their original estimates whenever they are significant. The estimates of the Maastricht dummy remain close to their baseline values and always retain their significance.

The trend increase in openness to a large extent coincides with the trend increase in $\overline{OAD25_t}$ and in the cases of "Expanding only" and "Contracting only", openness takes over (part of) the role of from $\overline{OAD25_t}$. The other demographic variables, i.e. the country-specific deviation from the OECD average and the changes in the projected old-age dependency ratio, are never significant. Interesting are the significant coefficients of the short-term interest rate and inflation rate. Lower interest rates and inflation tend to increase the probability of an "Expanding only" regime. Lower interest rates reduce debt-servicing costs, increasing room for more generosity in the pension system and making it cheaper to borrow in order to finance extra generosity. Lower inflation reduces the nominal interest rate, thus leading to the same effect. This is supported by the fact that the ex-post real interest rate, i.e. the nominal interest rate minus inflation, is not significant when it is included instead (not reported in Table D.1).

4.2.5 Controlling for crises

In our next robustness test, we investigate whether, in line with the crisis-induced reform hypothesis (see, for example, Drazen and Grilli, 1993, Tommasi and Velasco, 1996, Rodrik, 1996, and Drazen and Easterly, 2001), controlling for a crisis affects our baseline estimates. Based on the data in Laeven and Valencia (2018) we define a dummy $CRISIS_{it}$ that takes a value of one in all the cases in which they identify the occurrence of a banking crisis, a currency crisis or a sovereign debt crisis, as well as when a sovereign debt restructuring takes place. The number of instances a crisis coincides with one of the reform regimes is reasonably limited.[21] The estimates are found in the last row of Table D.1. The crisis dummy is only significant (at the 10% level) for the "Contracting and expanding" regime. For the other regimes, the significance of the other variables is unaffected. We also estimated the model with a dummy for each type of crisis. These dummies are never significant, while the baseline estimates always remain unchanged.

[21] The exact numbers of coincidences are 20 for $CRISIS_{it}$ and "Expanding only", 19 for $CRISIS_{it}$ and "Contracting only", and 19 for $CRISIS_{it}$ and "Contracting and expanding".

4.2.6 The role of political variables

Verbic and Spruk (2019) point to the potential roles of elections and the ideological leaning of the government.[22] This subsection tries to address the roles of these and other political factors. We do this by each time expanding the baseline regression with political variables and checking whether these are significant or not and whether the baseline coefficient estimates have undergone material changes. The set of political variables is chosen to cover a wide range of different aspects of the political environment. The political variables, obtained from Armingeon *et al.* (2018a, 2018b),[23] are (1) $CABIND_{it}$, which weighs the seats in the cabinet held by right-wing, centrist and left-wing parties: a higher value means a shift in seats from right- to left-wing parties; (2) $PARIND_{it}$, which captures the average political color of the parties in parliament: a higher value implies a more left-wing average orientation of the parliament; (3) $GOVCHANGE_{it}$, the number of government changes in a year; (4) $GOVNEW_{it}$, a dummy for whether the government is new; and (5) $ELECYEAR_{it}$, a dummy for an election year for the national parliament (lower house). To take account of the potential role of political factors in shaping legislation that needs some time to materialize, we include, besides the contemporaneous values, the first lags of all the political variables. For the election dummy, we also include the first lead, because many elections take place at regular points in time and, hence, are foreseen.

Table D.2 in online Appendix D report the estimates. Significant baseline variables always remain significant. The political variables play a rather limited role, essentially confirming what Duval et al. (2018) find for labor and product market reforms in advanced economies. The political color plays some role for the “Contracting only” regime: in line with what one might intuitively expect, a more left-wing government or parliament lowers the likelihood of a “Contracting only” regime. A general election in the next period reduces the chance of a “Contracting only” regime. In many instances, such an election is foreseen rather long before it takes place. Knowing that it is up for re-election, a government may be reluctant to take unpopular contracting measures. The current and lagged coefficients of the political color indicators offset each other for the “Contracting only” regime. A new government raises the likelihood of the “Contracting and expanding” regime, possibly because its mandate is at its strongest when it starts its term and its short-sightedness is at its lowest, because new elections are relatively far away.

We also added interactions of the political variables with the business cycle variables and the Maastricht dummy (see online Appendix E). On a few occasions, these interactions are significant,

[22] Cukierman and Tommasi (1998) argue that left-wing governments might be better placed at convincing the population of the long-run benefit of market-oriented reforms.
[23] For their construction see http://www.cpds-data.org/images/Update2018/Codebook-CPDS-1960-2016-Update-2018.pdf.

but the pattern of those interactions is not systematic and, therefore, difficult to interpret. Crucially, however, the coefficients on the (non-interacted) business cycle variables and their significance remain virtually the same as under the baseline. This is also the case for the demographic variables and the Maastricht dummy. Hence, our main findings continue to hold.

4.2.7 Additional robustness analysis and instrumental variables

We explore a number of additional variants for which results are reported in online Appendices D and E. First, replacing all projected old-age dependency ratio variables by their current values, i.e. replacing $\left(\overline{OAD25_t}, OADDEV25_{it}, \Delta OAD25_{it}\right)$ by $(\overline{OAD_t}, OADDEV_{it}, \Delta OAD_{it})$, where ΔOAD_{it} is the change in the current old-age dependency ratio between the years *t*-1 and *t*, always leaves the significance or insignificance of the business cycle variables unchanged (Table D.3 in online Appendix D), thus confirming the importance of the business cycle for pension reform decisions. Significance of the Maastricht dummy is also unchanged. Only in case of the "Contracting and expanding" regime, does $\overline{OAD_t}$ loses its significance. However, given the conceptual preference for $\overline{OAD25_t}$, we see no reason for changing the baseline.

Second, if we restrict the coefficients of the cross-country average old-age dependency ratio $\overline{OAD25_t}$ and the country deviation from this average, $OADDEV25_{it}$ to be identical, we effectively have the old-age dependency ratio $OAD25_{it}$ itself as explanatory variable. This yields essentially the same estimates as under the baseline (see Table E.4 in online Appendix E). The advantage of our baseline approach is that the split into average and deviation from that average reveals that the explanatory power comes from the former instead of the latter.

Third, we leave out the demographic variables with insignificant coefficient estimates, that is, the country deviation part, $OADDEV25_{it}$, and the first difference $\Delta OAD25_{it}$. The remaining estimates are virtually the same as for the baseline (Table E.5 in online Appendix E).

Fourth, we estimate the baseline model replacing the demographic variables with country-specific time trends. The coefficient estimates of the remaining baseline variables are unchanged (Table E.6 in online Appendix E).

Fifth, estimating the model as a conditional logit model also does not affect any of the results (Table D.4 in online Appendix D).

Sixth, we estimate a linear probability instead of logit model, where the probability of a reform regime is a linear function of the explanatory variables. Concretely, we estimate $h_{it,r} = \alpha_{0i,r} + \alpha_r' BASEVAR_{it} + \varepsilon_{it}$, where $h_{it,r}$ is one if regime r materializes and zero, otherwise. The regression fit is then interpreted as the probability that regime r will occur for given values of the explanatory variables. Obviously, this cannot be the theoretically correct model, because it may predict regime probabilities outside the zero-one range. Nevertheless, it is useful to see if the model estimates support the conclusions of the baseline model estimation, because the non-linearity of the logit specification and the relative sparsity of the reforms imply that the reform observations tend to be concentrated in the more non-linear part of the logit specification. We find that for all three regimes the baseline variables retain their significance (insignificance) if they were significant (insignificant) before (Table D.5 in online Appendix D), except for the Maastricht dummy for the "Contracting only" regime. The magnitudes of the effects, however, remain closely in line with those for the baseline logit regression. Hence, overall the linear specification supports the results from the baseline specification. The linear probability specification also makes it easy to do an instrumental variables estimation. As instruments we use the first lag of the explanatory variables. The estimates are found in Table D.6 in online Appendix D. The results remain close to the non-IV results.

Finally, to explore whether there is prima facie evidence that migration affects the results, we rank countries in terms of average gross migration flows during the sample period and split the sample into the top-5 subsample of migration countries (i.e., Australia, Canada, Ireland, Luxemburg and Switzerland) and the subsample of the remaining countries. The estimates, reported in Tables E.7 and E.8 in online Appendix E, exhibit no qualitative differences between the two groups of countries. Only the Maastricht dummy loses significance for the high migration countries, which is not unexpected since only two of these countries signed the Maastricht Treaty, so this subsample suffers from a lack of observations. In fact, if anything, the role of the business cycle in explaining reform regimes is quantitatively stronger for the high migration group.

4.3 Distinguishing broad versus specific groups

In this subsection we distinguish reform regimes according to whether they affect broad groups in society ("Many") or only specific groups ("Not many"). To this end, we estimate per regime type "Expanding only", "Contracting only" and "Contracting and expanding" a multinomial logit model, in which we allow the variable coefficients to differ between "Many" and "Not many". Table 11 reports the estimates for the baseline specification. For the sake of space, we only present the estimated coefficients, not the average marginal effects. Obviously, the number of observations per combination of regime type and width of group affected is substantially smaller than the number of observations per regime type only. This is in particular the case for "Not many". Still we observe that our main result, the role of the state of the business cycle is preserved for each group, although it appears to be more significant for "Many": growth has a significantly positive effect on "Expanding only" for both "Many" and "Not many". Growth has a highly-significantly negative effect on "Contracting only" in the case of "Many". For both "Many" and "Not many", the unemployment rate exerts a positive effect on "Contracting only". The deficit is highly-significantly positive for the combination "Contracting and expanding" and "Many", while it is insignificant for the combination

“Contracting and expanding” and “Not many”. There are at least two plausible explanations for the seemingly larger role of the state of the business cycle for “Many” compared to “Not many”. One concerns the smaller sub-sample size for “Not many”, as indicated in Table 4. A second reason may be that demographic considerations are more important when broad groups are affected, for example because of the future budgetary consequences (see next section).

An overall test whether the coefficients of the business cycle variables $(GROWTH_{it}, DEF_{it}, UNEMPL_{it})'$ are equal for “Many” and “Not many” within each regime is not rejected for the “Contracting only” regime and the “Contracting and expanding regime” as the Wald tests in the table show; this supports the homogeneity imposed in our baseline model. For the “Expanding only” regime, equivalence is just rejected at the 10% level, but this is mostly driven by the coefficients on growth in the 1970s. The estimates for the Maastricht dummy are in line with the baseline and quite comparable between “Many” and “Not many”. The projected future old-age dependency ratio exerts a highly-significant negative effect on “Expanding only” for both “Many” and “Not many”, but becomes insignificant for the other two regimes in the latter case, while it stays highly significant for “Many”.

5 A conceptual framework and interpretation

In our empirical analysis we found no effect of *changes* in the future or current old-age dependency ratio on the timing of pension reform measures. By contrast, the cyclical state of the economy, broadly defined, has a statistically robust significant effect on reform measures. This section discusses potential explanations for this combination of findings.

5.1 Fixed costs of implementing reform measures

This subsection discusses a conceptual framework that can explain our main empirical findings. A formal theoretical model with a minimum set of features that yields these outcomes is found in Beetsma and Romp (2019).

It builds on four key assumptions. First, the government trades off the individual pension benefit level and the level of public consumption, which both yield a positive marginal utility to the government. The government takes future utility into account, but may be voted out of office, so it has a quite high rate of time preference.

Second, business cycle fluctuations are a high-frequency process and the government’s endowment fluctuates with the business cycle. Hence, during a recession, there is more budgetary pressure on the government, forcing it to reduce both the individual pension benefit and other public spending. Total spending on the pension system obviously increases with the old-age dependency ratio.[24]

Third, the main forces underlying the demographic process – fertility and mortality – change slowly over time, which makes demographic fluctuations relatively predictable in the short to medium term. The implication is that it is safe to assume for a government that one or more business cycle fluctuations occur before a demographic shock actually affects the old-age dependency ratio.

Finally, changing the individual pension benefit comes at a fixed utility cost for the government. This cost may capture the loss of political capital from taking reform measures that may not be popular with (large groups of) voters. The fixed cost may differ across countries, thereby explaining why reform measures are more frequent in some countries than in others. The fixed cost may also differ between expanding and contracting measures, thereby helping to explain differences in the frequency of observing expanding versus contracting measures. The cost is likely higher for contracting measures. Indeed, we rarely observe people taking to the streets, because the welfare system is expanded. Theoretically, those who lose from contracting measures can be relatively easily identified and may have an interest in organizing resistance, while the costs of expanding measures will be typically spread out over broad groups of tax payers for whom the net benefit of organizing resistance is relatively low and who are likely unable to attribute changes in taxes to specific purposes.[25] Even so, there is likely some political cost of expanding measures as tax payments are unevenly distributed across the voting population and opposition parties may try to raise awareness of the cost of such measures. Each period the government chooses the optimal pension benefit, taking into account the state of the business cycle, the current and future demographic situation, and the costs of deviating from the current pension benefit. This is comparable to a standard optimal pricing problem with menu costs, but with one additional complexity: the government has additional information concerning the next period's demographic situation, since current shocks to fertility and mortality only affect future old-age dependency ratios.

Under these four assumptions, reform measures commonly coincide with business cycle fluctuations, but are seemingly unrelated to (projected) demographic changes. To illustrate, consider a situation where the government observes a demographic shock that will – in due time – render the pension system unsustainable. The current economic and demographic conditions, however, are such that the government would prefer to keep the current individual pension payout unchanged,

---

[24] In reality, pension benefits are a rather limited share of aggregate public spending and changes in pension generosity tend to be incremental or phased in over a longer period. Hence, the theoretical framework abstracts from feedback effects of changes in pension benefits on the cyclical state of the economy.

[25] Von Hagen and Harden (1995) demonstrate the tendency for governments to overspend in the context of common-pool models in which the benefits from spending are concentrated, while the cost in terms of taxes is shared by everybody.

provided it could ignore the future. That is, current instantaneous utility is maximized by keeping the individual pension payout at its current level. The government has two options: change the pension benefit now to make the pension system future proof or keep it the same now and change it in the future. In this situation, postponing the change in the pension benefit is clearly beneficial, since this has three advantages over the first option. First, the government postpones paying the fixed cost to the (heavily) discounted future. Second, the government does not have to set a pension benefit that differs from the current (close to) optimal level, thereby lowering its current instantaneous utility. Third, in the next period, the government can, if it decides to change the pension benefit, choose a benefit level that is optimal from that period onwards, taking new information into account. This illustrates why projected demographic shocks do not coincide with pension reforms.

If, on the other hand, the economy is hit by a positive business cycle shock, then the government wants to share the shock between the individual pension benefit and public consumption. The government's utility gains of changing the individual pension benefit (as part of the trade-off against changing other public spending) are immediate, not discounted. If the shock is large enough, then those immediate gains are sufficient to offset the fixed adjustment costs and the government will act immediately, regardless of the demographic forecasts. Analogously, in the case of a sufficiently large negative business cycle shock, will the government reduce the individual pension benefit. This can explain why business cycle fluctuations do show up as significant determinants of pension reform measures in our empirical analysis.

Our assumptions are realistic. Annual changes in the *current* demographic conditions are typically small (e.g., Preston *et al.*, 2000), so the instantaneous utility gains from resetting the benefit in response to such changes are small. Second, actual changes to old-age dependency ratios should have been anticipated and incorporated in an earlier change of the pension benefit. Hence, the gain from changing the benefit again after a demographic shock would generally be insufficient to overcome the fixed adjustment cost. The cumulative effect of small, predictable, changes in the current demographic conditions could at some point trigger a reform, but business cycle fluctuations take place at a higher frequency. Well before the cumulative effect of changes in current demographic conditions has become large enough to induce a reset of the benefit, likely a recession or boom will already have pre-empted this reset.

Changes in the (long-term) *projected* demographic conditions are unexpected, but merely change instantaneous utility in the distant future. Given the generally high effective rate of time preference of the current government, such change is unlikely to trigger current reform measures. Further, even if the government foresees sustainability problems due to adverse ageing shocks, it can safely postpone a reform since it is highly likely that a sufficiently-large business cycle fluctuation will trigger a reform well before sustainability issues have had a chance to materialize.

Finally, increasing public awareness of the future ageing costs may cause the fixed cost of expansionary measures to increase and that of contractionary measures to fall with the OECD-wide projection of the old-age dependency ratio. Given the state of the business cycle, it becomes (politically) more difficult to introduce expansionary measures and easier to introduce contractionary measures. The logic of our conceptual framework would then predict both the observed positive relationship between the frequency of contracting measures and the OECD-wide projected old-age dependency ratio and the negative relationship between the latter and the frequency of expanding measures.

5.2 Other explanations

Complementary explanations may contribute in rationalizing our findings. First, while implementing contracting measures is politically costly in general, being in a recession may increase the public's understanding of the need to introduce unpleasant measures. Conversely, if the economy is in a boom, public pressure may arise to increase welfare spending and have everyone share in the increase in resources. Not doing so may become politically costly. This may be enforced by a "this time is different syndrome" (Reinhart and Rogoff, 2011), where politicians and the general public may be over-optimistic about future growth.

Second, demographic projections could dictate the need for contractionary reform, but a (partly self-interested) government may choose policy trading off the effect on its re-election probability and future sustainability. Ceteris paribus, contractionary measures would make its re-election less likely, while the opposite holds for expanding measures. If the economy is in a boom, the government may be tempted to increase the generosity of the current system to buy re-election. However, if the country is in a recession, the government realizes that its re-election chances are already likely to be low and reducing pensions would not harm these much further, so in trading off its own interest and that of the current and future population at large, it focuses on what is sustainable in the long run and adopts the necessary contractionary reform.[26]

A final explanation for the lack of response to demographic shocks could be the electoral weight of the elderly. They are typically politically more active than the young (and certainly more active than future generations), which could allow them to block contractionary pension reforms.

[26] We thank an anonymous referee for suggesting this possible explanation. This mechanism connects to the political business cycle literature – for an overview, see Drazen (2000) and Dubois (2016). However, in the described mechanism, the politician's choices are driven by the business cycle, rather than the politician trying to manipulate the business cycle through its choices.

All these mechanisms can be viewed as complementary to that laid out in the previous subsection and the first two would in fact strengthen the role of the business cycle in explaining the likelihood of reform measures.[27]

6 Concluding remarks and policy implications

In this paper, we empirically explored the determinants of the timing of pension reform measures. To this end we used a narrative approach to construct a unique and comprehensive dataset of pension reform measures for a broad set of OECD countries since the start of the seventies. Reform measures are categorized as either expansionary or contractionary. The determinants that we considered comprised demographic, economic, budgetary, political and crisis variables, as well as being subject to the Maastricht Treaty. We tried to explain decisions on reform measures on the basis of information available at the time when reform measures were legislated.

We found that the OECD average projected old-age dependency ratio was a highly significant determinant of the trend reduction in expansionary and trend increase in contractionary measures. However, country-specific deviations from the OECD average projection and country-specific changes in the projected old-age dependency ratio had no significant explanatory power. By contrast, we found strong evidence that the current state of the business cycle, broadly defined, is an important driver of both expansionary and contractionary reform measures. Our interpretation is that while general (projected) demographic trends are responsible for the accumulating pressure to reform pension arrangements, it is the state of the business cycle that determines the timing of reform measures. We presented a simple conceptual framework with an adjustment cost of adopting reform measures that can account for both the reform responsiveness to the business cycle and its non-responsiveness to changes in the current demography and demographic forecasts.

Our analysis yields some potentially interesting policy implications. International institutions, like the OECD, the World Bank and the European Commission, advise countries to reform their pension arrangements so as to enhance their financial sustainability in anticipation of the ageing of their populations. This way the budgetary costs associated with ageing can be spread over the subsequent cohorts and space can be retained for the future provision of public services. Our estimates of the trend effects of changes in projected old-age dependency ratios are consistent with the fact that the advices of these international institutions combined with increasing national public and political awareness indeed stimulate reform. A further strengthening of the role of these

[27] Explicitly testing for the role of the first two mechanisms with the political variables at hand is difficult, since the mechanisms cut across political color, whether the government has changed or whether it is new.

institutions in the policy debate would be beneficial, because all advanced economies are confronted with population ageing. Hence, these institutions can compare different approaches across countries in dealing with ageing. In particular, they can highlight existing best practices and help countries in copying successful reforms.

Our finding that contractionary reforms occur disproportionally in periods of economic stress suggests that the cost-benefit trade-off in taking such measures is most favorable under those circumstances. However, this also implies that these measures come at the moment when they are least welcome to those who are directly affected by them.[28] From this perspective, it would generally be desirable to sever the link between the state of the business cycle and the moment of adoption of pension reform measures. A practical solution is to introduce some form of automaticity in sustainability-enhancing measures. An example is the automatic link between changes in life expectancy projections and adjustment of the retirement age for the public pension introduced by the Netherlands in 2012. Not only is the direct link with the state of the business cycle broken. The automaticity also silences discussion each time the retirement is raised, because the relationship between life expectancy and retirement age is enshrined in the law and the public gets accustomed to the idea of a prolongation of its working life. This also facilitates planning for retirement, while the uncertainty about the retirement age keeping up with increases in life expectancy shrinks. Obviously, most of the political cost is shifted to the moment when the automatic link is established. A way to ameliorate the "upfront cost" is to plan the first increase in retirement age sufficiently far in advance, thus taking away the resistance from those who are soon to retire and are likely most vocal about the "expropriation" of their rights, and exploiting the time discounting or even myopia of the younger cohorts vis-à-vis the measures.

The desirability to sever the link with the state of the business cycle extends to measures to enhance pension generosity. Motivated by the idea that everyone should share in the spoils, an economic boom generates pressure to expand social security, which effectively amounts to a reduction in the reform cost alluded to in the previous section, at a moment when the marginal benefit of such an expansion is relatively low. Such an expansion is particularly questionable when the underlying demographic trend requires an even larger offsetting contraction in the future. Appropriately-designed fiscal constraints may help to discourage such "unwarranted" expansion during boom periods. The EU's "Stability and Growth Pact" (SGP) requires countries to maintain a structurally-balanced public budget or, as long as they have not reached this position, to improve

[28] Obviously, this depends on the type of contractionary measure. A reduction in pension benefits has a direct effect on disposable income and we would expect this to hurt more during a recession than an increase in the retirement age.

their structural balance at a sufficient speed.[29] Expansion of public pension generosity would typically lead to a deterioration of the structural balance and would thus be discouraged by the rules.

However, the SGP is frequently violated and enforcement has proven difficult (e.g. see European Fiscal Board, 2018, 2019). An alternative is to enshrine appropriate fiscal restrictions in domestic high-level legislation and have these restrictions monitored by so-called national independent fiscal institutions (IFIs) that have been set up in the past decade in many countries. Whether or not this "legal route" is feasible, IFIs could be mandated to directly monitor the sustainability of public pension provisions as part of their role in monitoring the sustainability of the public finances.

References

Abiad, A. and A. Mody (2005). Financial Reforms: What Shakes it? What Shapes it? *American Economic Review* 95, 1, 66-88.

D'Amato, M. and V. Galasso (2010). Political Intergenerational Risk Sharing, *Journal of Public Economics* 94, 9-10, 628-637.

Armingeon, K., Wenger, V., Wiedemeier, F., Isler, C., Knöpfel, L., Weisstanner, D. and S. Engler (2018a), *Comparative Political Data Set 1960-2016*. Bern: Institute of Political Science, University of Berne.

Armingeon, K., Wenger, V., Wiedemeier, F., Isler, C., Knöpfel, L. and D. Weisstanner (2018b), *Supplement to the Comparative Political Data Set – Government Composition 1960-2016*. Bern: Institute of Political Science, University of Berne.

Beetsma, R. and X. Debrun (2007). The New Stability and Growth Pact: A First Assessment, *European Economic Review* 51, 2, 453-478.

Beetsma, R. and W. Romp (2019). Formalization of conceptual framework and implementation of formal framework, *Mimeo*, University of Amsterdam.

Beetsma, R. and H. Uhlig (1999). An Analysis of the Stability and Growth Pact, *Economic Journal* 109, 458, 546-571.

Bertola, G. and T. Boeri (2002). EMU Labour Markets Two Years on: Microeconomic Tensions and Institutional Evolution, in M. Buti and A. Sapir (eds.), *EMU and Economic Policy in Europe: The Challenges of the Early Years*, Edward Elgar, Aldershot.

Blinder, S. and A. Krueger (2004). What Does the Public Know About Economic Policy, And How Does It Know It? *NBER Working Paper*, No. 10787.

[29] The structural balance is the cyclically-adjusted balance corrected for one-off elements.

Boeri, T. and G. Tabellini (2012). Does Information Increase Political Support for Pension Reform? *Public Choice* 150, 1-2, 327-362.

Bonfiglioli, A. and G. Gancia (2015). Economic Uncertainty and Structural Reforms, *CEPR Discussion Paper*, No.10937.

Brooks, S. (2007a), When does diffusion matter? Explaining the spread of structural pension reforms across nations. *Journal of Politics*, 69, 3, 701-715.

Brooks, S. (2007b). Globalization and Pension Reforms in Latin America. *Latin American Politics and Society* 49, 4, 31 – 62.

Campos, N., Hsiao, C. and J. Nugent (2010). Crises, what Crises? New Evidence on the Relative Roles of Political and Economic Crises in Begetting Reforms. *Journal of Development Studies* 46, 10, 1670-1691.

Castanheira, M., Galasso, V., Carcillo, S., Nicoletti, G., Perotti, E. and L. Tsyganok (2006). How to Gain Political Support for Reforms, in Boeri, T., Castanheira, M., Faini, R. and V. Galasso (eds.), *Structural Reform without Prejudices*, Oxford University Press.

Chamberlain, G., (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies* 47, 225-238.

Cooley, T.F. and J. Soares (1999). A Positive Theory of Social Security Based on Reputation, *Journal of Political Economy* 107, 1, 135-160.

Cukierman, A. and M. Tommasi (1998). When Does it Take a Nixon to Go to China? *American Economic Review* 88, 1, 180-97.

Drazen, A. (2000). The Political Business Cycle after 25 years. *NBER Macroeconomics Annual* 15, 75–117.

Drazen, A. and W. Easterly (2001). Do Crises Induce Reform? Simple Empirical Tests of Conventional Wisdom, *Economics and Politics*, 13, 2, 129-157.

Drazen, A. and V. Grilli (1993). The Benefits of Crises for Economic Reforms, *American Economic Review* 83, 3, 598-607.

Dubois, E. (2016). Political Business Cycles 40 years after Nordhaus. *Public Choice* 166, 1, 235–259.

Duval, R., Furceri, D. and J. Miethe (2018). The Needle in the Haystack: What Drives Labor and Product Market Reforms in Advanced Countries? *IMF Working Paper*, WP/18/101.

European Commission, (2014). *AMECO Database*. Retrieved from http://ec.europa.eu/economy_finance/db_indicators/ameco/index_en.htm.

European Fiscal Board (2018, 2019), *Annual Report*, Brussels, https://ec.europa.eu/european-fiscal-board.

Galasso, V. and P. Profeta (2002). The Political Economy of Social security: a Survey, *European Journal of Political Economy* 18, 1, 1-29.

Giuliano, P., Prachi, M. and A. Spilimbergo (2013), Democracy and Reforms: Evidence from a New Dataset, *American Economic Journal: Macroeconomics*, 5, 4, 179-204.

Gonzales-Eiras, M. and D. Niepelt (2008). The Future of Social Security, *Journal of Monetary Economics* 55, 197-218.

Gordon, M.S. (1988), *Social Security Policies in Industrial Countries: A Comparative Analysis*, Cambridge University Press, Cambridge.

Gruber, J. and D. A. Wise (2009). *Social Security Programs and Retirement around the World: Fiscal Implications of Reform*, University of Chicago Press, Chicago, IL.

Hagen, J. von and I.J. Harden (1995). Budget Processes and Commitment to Fiscal Discipline, *European Economic Review*, 39, 3-4, 771-779.

ILO (2019). *International Labour Organization: NATLEX: Database of National Labour, Social Security and Related Human Rights Legislations*. http://www.ilo.org/dyn/natlex, access date: March 1 and March 20, 2019.

IMF (2019). *World Economic and Financial Surveys: World Economic Outlook Database*. Retrieved from https://www.imf.org/external/pubs/ft/weo/2019/01/weodata/index.aspx.

ISSA (2019). https://www.issa.int/en_GB/country-profiles, access date: March 8, 2019.

James, E. and S. Brooks (2001). The Political Economy of Structural Pension Reform, in Holzmann, R. and J. Stiglitz (eds.), *New Ideas About Old Age Security*, World Bank, Washington DC.

LABREF (2019). https://webgate.ec.europa.eu/labref/public/, access date: April 17, 2019.

Laeven, L. and F. Valencia, (2018). Systemic Banking Crises Revisited, *IMF Working Paper*, No. WP/18/206.

Leibrecht, M. and J.H. Fong (2017). Drivers of Market-based Pension Reforms: Crises and Globalisation, *Discussion Paper*, No. ICM-2017-05, Henley Business School.

Mahmalat, M. and D. Curran (2018). Do Crises Induce Reform? A Critical Review of Conception, Methodology and Empirical Evidence of the 'Crisis Hypothesis', *Journal of Economic Surveys* 32, 3, 613-648.

Madrid, R. (2002). The Politics and Economics of Pension Privatization in Latin America. *Latin American Research Review* 37, 2.

OECD (2012, 2014). *Pension Outlook*. OECD Publishing, Paris.

OECD (2007, 2009, 2011, 2013, 2015, 2017). *Pensions at a Glance*. OECD Publishing, Paris.

Orenstein, M. (2005). The New Pension Reform as Global Policy. *Global Social Policy* 5, 2, 195-202.

Orenstein, M. (2013). Pension Privatization: Evolution of a Paradigm. *Governance: An International Journey of Policy, Administration, and Institutions* 26, 2, 259-281.

Persson, T. and G. Tabellini (2000). *Political Economics: Explaining Economic Policy*, MIT Press, Cambridge, MA.

Preston, S., Heuveline, P., and M. Guillot (2000). *Demography: Measuring and Modeling Population Processes*, John Wiley and Sons, Hoboken, NJ.

Razin, A., Sadka, E. and P. Swagel (2002). The Aging Population and the Size of the Welfare State, *Journal of Political Economy* 110, 4, 900-918.

Ranciere, R. and A. Tornell (2015). Why Do Reforms Occur in Crises Times? *Mimeo*, IMF and UCLA.

Reinhart, C.M. and K.S. Rogoff (2009). *This Time Is Different: Eight Centuries of Financial Folly*, Princeton University Press, Princeton, NJ.

Rodrik, D. (1996). Understanding Economic Policy Reform. *Journal of Economic Literature* 34, 9-41.

Swagel, P. (2015), Legal, Political, and Institutional Constraints on the Financial Crisis Policy Response, *Journal of Economic Perspectives* 29, 2, 107-22.

Tabellini, G. (2000). A Positive Theory of Social Security, *Scandinavian Journal of Economics* 102, 3, 523-545.

Thompson, W. (2009). *The Political Economy of Reform: Lessons from Pensions, Product Markets, and Labour Markets in Ten OECD Countries*, OECD Publishing, Paris.

Tommasi, M. and A. Velasco (1996). Where Are We in the Political Economy of Reform? *Journal of Policy Reform* 1, 187-238.

Tommasi, M. (2017). Crisis, Political Institutions, and Policy Reform: the Good, the Bad, and the Ugly, *Mimeo*, Universidad de San Andrés and Yale University.

United Nations (2017), World Population Prospects: The 2017 Revision, Methodology of the United Nations Population Estimates and Projections, *Working Paper*, No. ESA/P/WP.250. New York: United Nations.

Verbic, M. and R. Spruk (2019). Political Economy of Pension Reforms: An Empirical Investigation. *European Journal of Law and Economics* 47, 2, 171-232.

World Bank. (1994). *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*, World Bank, Washington, DC.

World Bank. (2019). *World Databank: World Development Indicators*. Retrieved from http://databank.worldbank.org/data/reports.aspx?source=2&Topic=3#.

Tables

Table 1: Number of (country, year) - combinations by category and reform regime

	1970 - 2017	1970 - 1993	1994 - 2017
Coverage (1a)	140	71	69
Generosity and adequacy (1b)	247	96	151
Expanding – other (1c)	10	1	9
Expanding (multiples equal one)	339	143	196
“Expanding only” regime	223	121	102
Fiscal sustainability (2a)	181	43	138
Work incentives (2b)	97	10	87
Contracting – other (2c)	5	0	5
Contracting (multiples equal one)	235	49	186
“Contracting only” regime	119	27	92
“Contracting and expanding” regime	116	22	94
Total (multiples equal one)	458	170	288

Note: “Expanding (multiples equal one)” reports the number of different (country, year) - combinations with one or more expansionary measures. Because there are (country, year) - combinations with both a “Coverage (1a)” and/or “Generosity and adequacy (1b)” and/or “Expanding – other (1c)” measure, the sum of “Coverage (1a)”, “Generosity and adequacy (1b)”, and “Expanding – other (1c)” exceeds “Expanding (multiples equal one)”. Analogous for “Contracting (multiples equal one)” and “Total (multiples equal one)”.

Table 2: Numbers of combinations of contracting and expanding measures

	Fiscal sustainability	Work incentives
Coverage	33	20
Generosity and Adequacy	70	42

Note: this table reports the number of country-year pairs of specific combinations of contracting and expanding measures.

Table 3: Unconditional frequency of reform regime and frequency conditional on regime in the previous year

	Expanding only (t-1)	Contracting only (t-1)	Contracting and expanding (t-1)	Unconditional
Expanding only	27%	15%	20%	20%
Contracting only	8%	16%	19%	11%
Contr. and exp.	9%	22%	12%	11%

Note: the first three data columns report the frequency of a regime given a specific regime in the previous year. The final column gives the unconditional frequency of a regime.

Table 4: Division of reform regimes into “Not many” and “Many”

	Total	Not many	Many
Expanding only	223	90 (40%)	133 (60%)
Contracting only	119	24 (20%)	95 (80%)
Contracting and expanding	116	11 (9%)	105 (91%)
Total	458	125 (27%)	333 (73%)

Table 5: Correlation matrix of business cycle variables

	$GROWTH_{it}$	DEF_{it}	$UNEMPL_{it}$
$GROWTH_{it}$	1		
DEF_{it}	-0.223	1	
$UNEMPL_{it}$	-0.153	0.437	1

Table 6: Logit estimations for the baseline regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and Expanding	
	Coeff.	Marg. eff.	Coeff.	Marg. eff.	Coeff.	Marg. eff.
$\overline{OAD25_t}$	-6.49***	-1.02***	4.77***	0.42***	4.66***	0.43***
	(1.65)	(0.26)	(1.72)	(0.15)	(1.75)	(0.16)
$OADDEV25_{it}$	2.04	0.32	0.091	0.0080	-0.68	-0.063
	(2.95)	(0.46)	(3.54)	(0.31)	(3.17)	(0.30)
$\Delta OAD25_{it}$	-2.51	-0.39	5.11	0.45	8.08	0.75
	(6.93)	(1.09)	(7.86)	(0.69)	(8.35)	(0.78)
$GROWTH_{it}$	15.7***	2.47***	-13.5***	-1.19***	-3.91	-0.37
	(4.38)	(0.67)	(4.78)	(0.42)	(4.63)	(0.43)
$70s \times GROWTH_{it}$	-3.68	-0.48				
	(5.94)	(0.79)				
DEF_{it}	-2.12	-0.33	1.86	0.16	8.79**	0.82**
	(3.08)	(0.48)	(3.82)	(0.34)	(3.64)	(0.34)
$UNEMPL_{it}$	-4.38	-0.69	17.3***	1.52***	1.06	0.099
	(3.73)	(0.59)	(4.98)	(0.44)	(4.16)	(0.39)
$MAASTRICHT_{it}$	0.84***	0.13***	0.91**	0.080**	0.91**	0.085**
	(0.28)	(0.044)	(0.44)	(0.039)	(0.41)	(0.038)
N	1081	1081	1034	1034	987	987
McFadden R2	0.076		0.17		0.13	

Notes: (i) The specification for "Expanding only" also includes an intercept term interacted with "$70s$", which is a dummy variable with value one for the years 1970 – 1979, and zero from 1980 until the end of our sample. The term is insignificant and for the sake of space not reported. For "Expanding only" the coefficient estimate and marginal effect of $GROWTH_{it}$ apply only to the period from 1980 onwards, while the coefficient estimate and marginal effect of $70s \times GROWTH_{it}$ apply only to the period 1970 – 1979. (ii) Standard errors are in brackets, *** denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. (iii) "N" is the number of observations. (iv) "Coeff." is coefficient estimate, "Marg. eff." is average marginal effect over all observations. (v) Country-fixed effects are included; no time-fixed effects are included.

Table 7: Including time-fixed effects in the baseline regression

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and Expanding	
	Coeff.	Marg. eff.	Coeff.	Marg. eff.	Coeff.	Marg. eff.
$\overline{OAD25_t}$	-5.77*	-0.85*	-3.45	-0.38	2.15	0.21
	(3.12)	(0.46)	(15.0)	(1.65)	(5.03)	(0.49)
$OADDEV25_{it}$	2.05	0.30	1.72	0.19	-0.61	-0.060
	(3.14)	(0.46)	(3.77)	(0.41)	(3.67)	(0.36)
$\Delta OAD25_{it}$	-1.60	-0.24	1.12	0.12	5.53	0.54
	(7.78)	(1.15)	(8.82)	(0.97)	(9.86)	(0.97)
$GROWTH_{it}$	17.0***	2.50***	-19.8***	-2.18***	-11.4	-1.11
	(5.19)	(0.76)	(7.09)	(0.77)	(7.21)	(0.70)
$70s \times GROWTH_{it}$	-4.09	-0.50				
	(7.60)	(0.91)				
DEF_{it}	-2.50	-0.37	1.61	0.18	10.8***	1.06***
	(3.37)	(0.50)	(4.57)	(0.50)	(4.14)	(0.40)
$UNEMPL_{it}$	-2.95	-0.43	13.1**	1.44**	3.93	0.39
	(4.11)	(0.61)	(5.50)	(0.60)	(4.94)	(0.48)
$MAASTRICHT_{it}$	0.73**	0.11**	0.88*	0.096*	0.24	0.024
	(0.34)	(0.050)	(0.51)	(0.056)	(0.54)	(0.053)
N	1081	1081	792	792	861	861
Wald test for all time fixed effects = 0:						
chi2	37.3		16.8		31.9	
Prob > chi2	0.72		0.98		0.71	

Notes: See notes to Table 6. Further, the time-fixed effect in 2010 has been normalized to zero to keep $\overline{OAD25_t}$ in the model. Although its coefficient is no longer meaningful, the advantage of keeping $\overline{OAD25_t}$ in is that, when testing the null hypothesis that the time-fixed effects are jointly zero, its explanatory power is exploited under the null, so that we test whether time-fixed effects are zero *after* accounting for $\overline{OAD25_t}$. Finally, the specification for "Expanding only" does not contain a dummy for the 70s. "Prob > chi2" is the p-value of the Wald test for that null hypothesis.

Table 8: Multinomial logit estimations with baseline specification

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and Expanding	
	Coeff.	Marg. eff.	Coeff.	Marg. eff.	Coeff.	Marg. eff.
$\overline{OAD25_t}$	-5.06***	-0.93***	3.60*	0.34**	3.91**	0.37**
	(1.73)	(0.26)	(1.92)	(0.15)	(1.98)	(0.16)
$OADDEV25_{it}$	3.54	0.51	1.41	0.046	0.71	-0.036
	(3.04)	(0.46)	(3.71)	(0.31)	(3.39)	(0.28)
$\Delta OAD25_{it}$	-0.16	-0.35	6.93	0.43	9.86	0.73
	(7.23)	(1.08)	(8.22)	(0.68)	(8.79)	(0.72)
$GROWTH_{it}$	13.2***	2.36***	-11.9**	-1.23***	-3.66	-0.37
	(4.51)	(0.67)	(5.27)	(0.44)	(5.38)	(0.43)
$70s \times GROWTH_{it}$	-4.00	-0.18	-27.9	-0.95	-10.5	-0.56
	(6.11)	(0.87)	(19.4)	(0.84)	(14.0)	(1.15)
DEF_{it}	0.14	-0.24	3.85	0.17	9.28**	0.73**
	(3.22)	(0.49)	(4.08)	(0.34)	(3.87)	(0.32)
$UNEMPL_{it}$	-2.07	-0.71	16.4***	1.40***	3.95	0.088
	(3.91)	(0.59)	(5.22)	(0.43)	(4.61)	(0.37)
$MAASTRICHT_{it}$	1.12***	0.12***	1.30***	0.070*	1.26***	0.063*
	(0.29)	(0.043)	(0.45)	(0.038)	(0.42)	(0.035)
N			1081			
McFadden R2			0.15			

Notes: See notes to Table 6. Further, country-fixed effects are included; no time-fixed effects are included.

Table 9: Including lags of the regime indicators

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and Expanding	
	Coeff.	Marg. eff.	Coeff.	Marg. eff.	Coeff.	Marg. eff.
$\overline{OAD25_t}$	-6.35***	-1.00***	5.07***	0.44***	4.77***	0.44***
	(1.68)	(0.27)	(1.77)	(0.15)	(1.80)	(0.16)
$OADDEV25_{it}$	2.08	0.33	0.44	0.039	-0.50	-0.045
	(2.96)	(0.46)	(3.59)	(0.31)	(3.16)	(0.29)
$\Delta OAD25_{it}$	-2.43	-0.38	3.84	0.34	10.1	0.93
	(6.93)	(1.09)	(7.95)	(0.69)	(8.51)	(0.78)
$GROWTH_{it}$	15.7***	2.47***	-13.2***	-1.15***	-4.42	-0.41
	(4.35)	(0.67)	(4.80)	(0.42)	(4.65)	(0.43)
$70s \times GROWTH_{it}$	-3.42	-0.44				
	(5.96)	(0.79)				
DEF_{it}	-2.14	-0.34	2.05	0.18	9.47**	0.87**
	(3.09)	(0.48)	(3.90)	(0.34)	(3.77)	(0.34)
$UNEMPL_{it}$	-4.20	-0.66	18.2***	1.59***	0.085	0.0078
	(3.79)	(0.60)	(5.03)	(0.44)	(4.23)	(0.39)
$MAASTRICHT_{it}$	0.82***	0.13***	0.96**	0.084**	0.96**	0.088**
	(0.28)	(0.044)	(0.44)	(0.039)	(0.42)	(0.038)
$EXP_{i,t-1}$	0.092	0.014	-0.32	-0.028	-0.19	-0.017
	(0.19)	(0.031)	(0.32)	(0.028)	(0.30)	(0.027)
$CON_{i,t-1}$	-0.27	-0.042	-0.58*	-0.050*	0.39	0.036
	(0.31)	(0.049)	(0.33)	(0.028)	(0.30)	(0.028)
$CONEXP_{i,t-1}$	0.16	0.025	0.038	0.0033	-0.87**	-0.079**
	(0.28)	(0.044)	(0.32)	(0.028)	(0.36)	(0.033)
N	1081		1034		987	
McFadden R2	0.078		0.17		0.14	

Notes: See notes to Table 6. Further, $EXP_{i,t-1}$ is a dummy for the occurrence of the "Expanding only" regime in period *t*-1, $CON_{i,t-1}$ is a dummy for the occurrence of the "Contracting only" regime in period *t*-1, and $CONEXP_{i,t-1}$ is a dummy for the occurrence of the "Contracting and expanding" regime in period *t*-1. Country-fixed effects are included; no time-fixed effects are included.

Table 10: Including averages of current and lagged business cycle variables

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and Expanding	
	Coeff.	Marg. eff.	Coeff.	Marg. eff.	Coeff.	Marg. eff.
$\overline{OAD25_t}$	-6.65***	-1.05***	4.36**	0.38**	4.56**	0.43**
	(1.68)	(0.27)	(1.75)	(0.15)	(1.78)	(0.17)
$OADDEV25_{it}$	2.02	0.32	0.037	0.0033	-0.50	-0.047
	(2.95)	(0.46)	(3.55)	(0.31)	(3.19)	(0.30)
$\Delta OAD25_{it}$	-2.41	-0.38	5.52	0.49	8.57	0.80
	(7.01)	(1.10)	(7.77)	(0.68)	(8.33)	(0.78)
$GROWTH_{i,t-1:t}$	16.8***	2.64***	-19.7***	-1.74***	-6.30	-0.59
	(4.99)	(0.77)	(6.03)	(0.53)	(5.71)	(0.53)
$70s \times GROWTH_{i,t-1:t}$	-0.27	-0.035				
	(7.64)	(1.00)				
$DEF_{i,t-1:t}$	-4.25	-0.67	1.96	0.17	9.91**	0.92**
	(3.31)	(0.52)	(4.17)	(0.37)	(3.98)	(0.37)
$UNEMPL_{i,t-1:t}$	-1.95	-0.31	14.6***	1.29***	-1.03	-0.096
	(3.75)	(0.59)	(5.03)	(0.44)	(4.35)	(0.41)
$MAASTRICHT_{i,t-1:t}$	0.83***	0.13***	0.91**	0.080**	0.93**	0.087**
	(0.28)	(0.044)	(0.44)	(0.039)	(0.41)	(0.038)
N	1081		1034		987	
McFadden R2	0.075		0.16		0.13	

Notes: See notes to Table 6. Averages of variables over the past and the current year are denoted by subscript "t-1:t". Country-fixed effects are included; no time-fixed effects are included.

Table 11: Multinomial logit baseline estimations distinguishing broad versus narrow groups

	(1)	(2)	(3)	(4)	(5)	(6)
Independent variables	Expanding only		Contracting only		Contracting and expanding	
	Not many	Many	Not many	Many	Not many	Many
$\overline{OAD25_t}$	-7.46***	-5.97***	1.46	5.86***	2.53	4.67**
	(2.54)	(2.01)	(3.51)	(1.94)	(5.37)	(1.84)
$OADDEV25_{it}$	1.80	2.30	2.42	-0.73	-10.9	0.050
	(5.00)	(3.50)	(7.88)	(3.87)	(12.0)	(3.33)
$\Delta OAD25_{it}$	5.30	-9.18	3.80	6.12	24.7	6.57
	(9.86)	(8.88)	(16.3)	(8.63)	(26.7)	(8.66)
$GROWTH_{it}$	11.2*	18.7***	0.71	-16.1***	-2.45	-3.92
	(6.11)	(5.24)	(10.9)	(5.19)	(13.2)	(4.87)
$70s \times GROWTH_{it}$	-14.6*	3.88				
	(8.47)	(7.77)				
DEF_{it}	-0.93	-3.62	11.0	1.06	-12.6	10.9***
	(4.39)	(3.91)	(8.99)	(4.18)	(12.9)	(3.82)
$UNEMPL_{it}$	0.99	-7.53	23.1**	15.4***	15.4	0.038
	(5.55)	(4.69)	(9.82)	(5.53)	(14.9)	(4.34)
$MAASTRICHT_{it}$	0.56	1.07***	0.75	0.93*	1.04	0.92**
	(0.41)	(0.35)	(0.82)	(0.51)	(1.35)	(0.43)
N	1081		1081		1081	
McFadden R2	0.094		0.19		0.18	
Wald test for equality of business cycle coefficients (of “Not many” and “Many”)						
chi2	7.82		3.37		3.34	
Prob > chi2	0.098		0.34		0.34	

Notes: See notes to Table 6. Further, country fixed effects are included; no time fixed effects are included.

Figures

Figure 1: Frequencies of the different reform regimes in each sample year

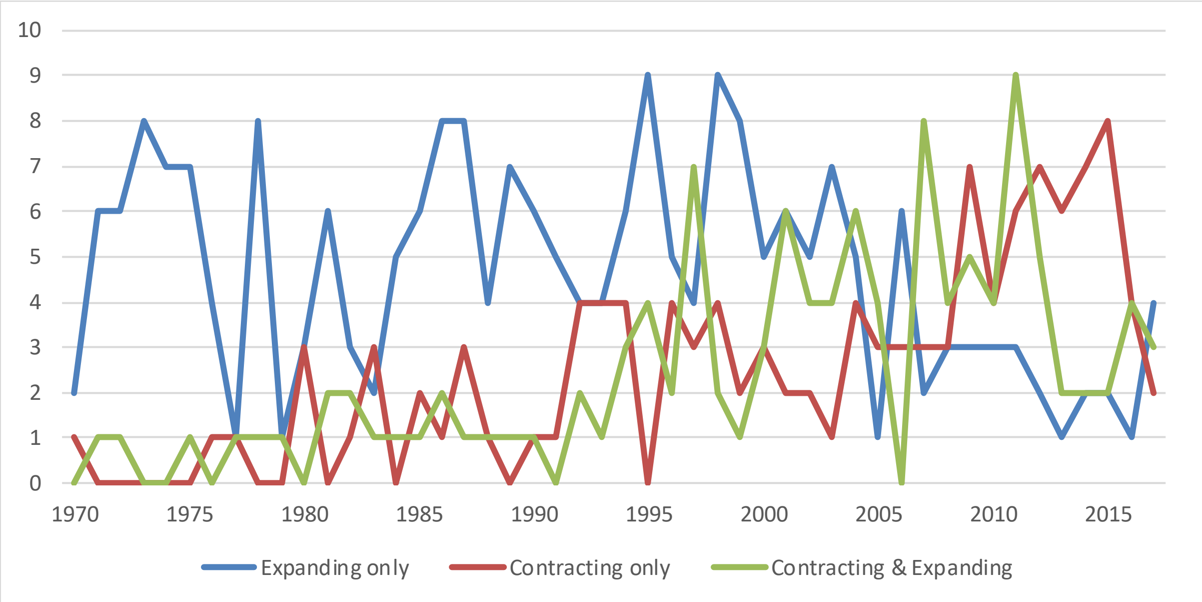


Figure 2: Frequencies of the different reform regimes in each country

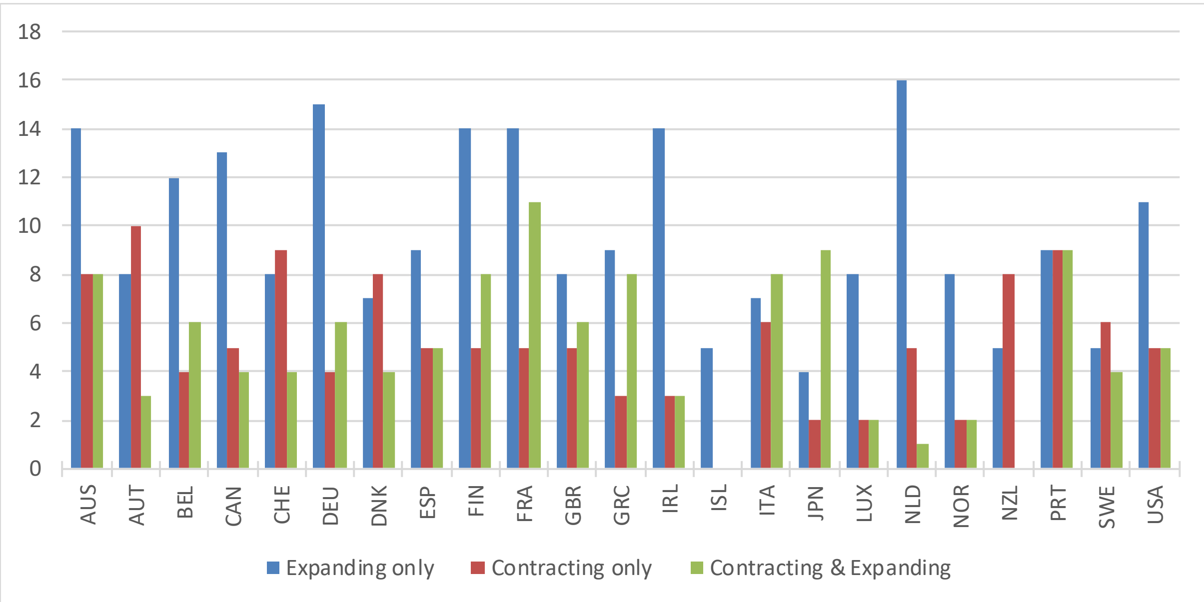
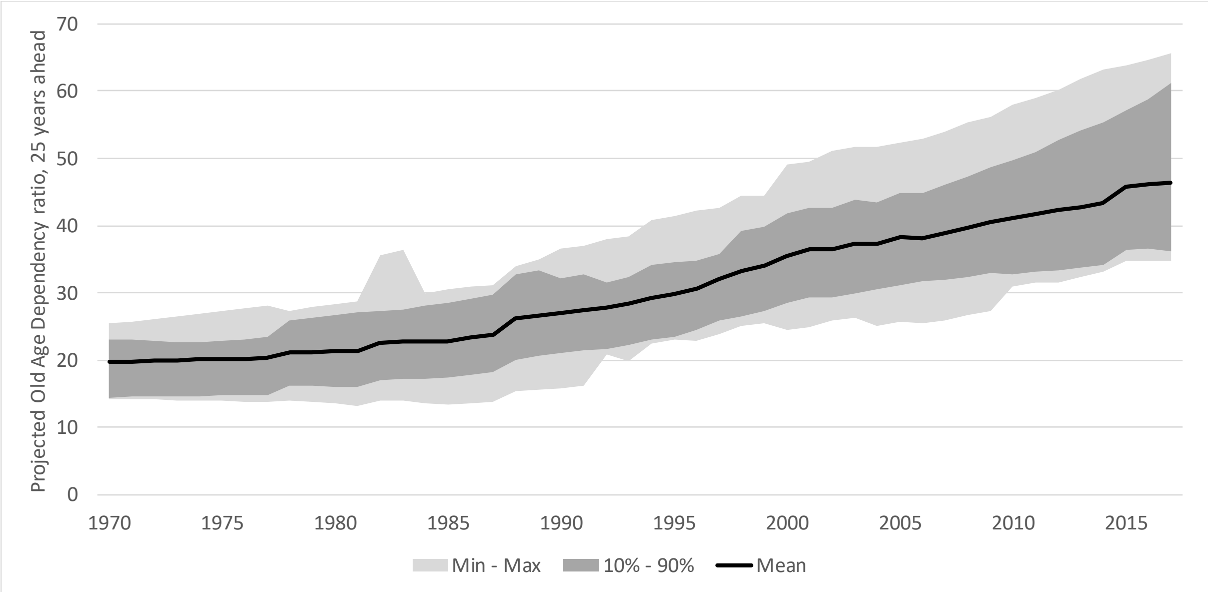


Figure 3: Twenty-five year ahead forecast of old-age dependency ratio.



Appendices

A. Details on the collection of the data on pension reform measures

Our goal is to construct a database based on narrative identification that categorises all reform measures that impact both the present value of retirement income and the government's intertemporal budget constraint. In our empirical analysis, we lump together all reform measures legislated in a given country in a given year and categorise them into a single reform regime. To do so, we prefer classifying reform measures individually, and then aggregate them into a country-year indicator. The advantage of the latter approach is that it is often easier (the relevant text is shorter) and more objective to categorise one reform measure at a time, than the combination of multiple reform measures. The drawback is that each reform measure must have a clear description. This is feasible for reform measures obtained from the ISSA database, the OECD publications, the LABREF database, and the ad-hoc sources. Unfortunately, for one database, NATLEX, this preferred approach is not feasible due to insufficient information, and we first have to group all reform measures in a country-year to get a better overview of the legislation passed in that country-year. Hence, for this database only, we categorise the reform measures by country-year instead of individually.

We classify all reforms, or groups of reforms in a country-year, into the following categories:

- "Unclear": reform measures that lack a clear description.
- "Clear": reform measures that feature a clear description. These can be divided into reform measures that are
  - "Not relevant": reform measures that have a clear description, but are not relevant for our analysis. An example is a reform that falls under the OECD's description of "administrative efficiency".
  - "Relevant": all other reforms, i.e. all reforms that affect the present value of retirement income and the government's intertemporal budget constraint.

The relevant reforms are classified along two dimensions: the first concerns their effect on the government's intertemporal budget constraint and the second concerns whether specific or broad groups in society are affected by the reform. As regards the effect on the government's intertemporal budget constraint, the subdivision is:

1a. "Expanding – Coverage": reforms that expand the coverage of the system.

1b. "Expanding – Generosity and adequacy": reforms that increase the benefits or lower the contributions.

1c. “Expanding – Other”: other measures that have a clear expanding nature.

2a. “Contracting – Fiscal sustainability”: reforms that increase the fiscal sustainability.

2b. “Contracting – Work incentives”: reforms that result in incentives to postpone retirement via work incentives.

2c. “Contracting – Other”: other reforms that have a clearly contracting nature.

3. Effect on government intertemporal budget constraint is unclear.

As regards the second dimension, that is, the width of the group affected, the subdivision is:

i. Broad groups, such as (almost) all females and/or males and/or elderly and/or working. We refer to this as “Many”.

ii. Narrow groups (e.g. only workers in specific sectors, only those born in a specific year). We refer to this as “Not many”.

iii. Affected groups unclear. We include these cases under “Not many”, based on the implicit assumption that the affected groups are insufficiently broad to justify being mentioned specifically.

The division in the first two groups allows us to distinguish between supposedly more and less substantial measures.

Our reforms come from four main sources: (1) the ILO’s NATLEX database, (2) the ISSA database on social security, (3) OECD publications on national pension systems, and (4) the LABREF database managed by the European Commission in cooperation with the Employment Committee (EMCO). Besides these four main databases, we consulted other, often country-specific publications on the history of pension reform. Below we discuss the data from each of the main databases.

<u>The NATLEX database</u>:

The ILO maintains its NATLEX database which contains all legislation related to labor law. For our analysis subject “15.02: Old age, invalidity and survivors’ benefit” is relevant. For a reason unknown to us, seafarers have a separate subject, “18.08 Seafarers – Social Security”. In total, these two subjects together contain 2,808 records that apply to our selection of OECD countries and the sample 1970-2017. The database contains an identifier unique for every reform measure (called ISN). Some reform measures apply to more than one country (but these are easily identified). Moreover, every ISN record contains a URL to a (slightly) more detailed description, sometimes links to original (national) legislation and often a more detailed description, called “citation” in NATLEX. The database is large, but unfortunately, not complete. The last Dutch reform in NATLEX, for example, is from 2005. It seems that some countries stopped submitting information about their legislation to NATLEX. This implies that we cannot rely on NATLEX as a sole source of information about legislation relevant for our analysis. Another limitation of the NATLEX database is that the description of legislation is often not very informative, sometimes even missing. For example, it may contain merely a table of content of the accepted legislation. Even when a legislation according to its description clearly affects benefits, it may be unclear whether the legislation is contracting or expanding. Those records that have a clear description are categorized into one of the above categories and directly included in our database. We can often link reform measures described in country-specific publications to records in the NATLEX database. Whenever this is possible, we include them as coming from NATLEX, together with the source of the additional information. In total, we include 322 NATLEX records directly.

For our regression analysis we merely need a classification at the country-year level, not the classification of every reform measure individually. Hence, instead of ignoring those reforms that we cannot classify individually, we combine all legislation passed in a single country-year in order to try and establish a clear classification for the combination of reform measures. This leads to 152 country-year combinations that provide a clear description and are relevant for our analysis.

<u>The ISSA database</u>:

The ISSA keeps track of the current regulations regarding social security and funded schemes. The ISSA Country Profiles cover national social security programs established by statute that insure individuals against interruption or loss of earnings resulting from old age, disability or survivorship; sickness and maternity; work injury or occupational disease; unemployment; and child raising. Much of the information in the Social Security Country Profiles is collected through surveys carried out by the ISSA in co-operation with the US Social Security Association. Data is also provided by numerous social security officials throughout the world. The ISSA claims to include reforms implemented from 1994 until 2018, but for our OECD countries the database only contains reforms from 1995 until (including) 2014. The ISSA database contains 1100 records that apply to our selection of OECD countries. Each record in the ISSA database consists of a country, year, unique “FicheNo”, short title, (long) description, legislation date and sources. The year is usually the year of implementation. Luckily, the footnotes also contain the year of legislation, which we use in our analysis. Many records describe reform measures, but some merely contain a description of a report that was released in that specific year. All records contain a sufficient description to classify them, but 822 of them are not relevant for our analysis – these are usually proposed measures that are not effectuated in legislation, reports discussed by the government, or reform measures in non-pension related fields such as health insurance.

<u>The OECD</u>

The OECD does not have a specific database, but includes in its regular publications on pensions tables of reform measures implemented in member countries. "Pensions at a Glance" is published every two years since 2005. The "OECD Pension Outlook" is published every two years since 2012. The publications that include a table with reforms measures are:

- Pensions at a Glance 2007, pp. 58-60: description of reform measures to national retirement income systems since 1990, but no data on year of legislation.
- Pensions at a Glance 2009, pp. 90-94: description of pension reform measures from 2004 to 2009, but no data on year of legislation.
- Pensions at a Glance 2013, pp. 27-40: details of pension reform measures enacted or implemented between January 2009 and September 2013, with year of legislation or implementation.
- Pensions at a Glance 2015, pp. 34-43: details of pension reform measures, September 2013 – September 2015, with year of legislation or implementation.
- Pensions at a Glance 2017, pp. 32-40: overview of pension reform measures decided between September 2015 and September 2017
- Pension Outlook 2012, Annex 1.A1, pp. 35-55: details of pension-reform measures legislated in the period September 2007-February 2012, by primary objective and with year of legislation.
- Pension Outlook 2014, Annex 2.A1, pp. 72-81: details of pension reform measures legislated in the period February 2012-September 2014, by primary objective and with year of legislation.

The other publications (in 2005, 2011, 2016, and 2018) do not contain tables with reforms. All publications together contain 615 reform measures in our selection of OECD countries.

In a first step we determined whether the reform measure is clear and relevant. This leaves 494 reform measures for which we need to determine the year of legislation. This turns out to be problematic for especially the early publications since we only know that reforms mentioned in those publications pre-date those publications. Luckily, 259 of these 494 are were easily matched to the extensively described reforms in the ISSA database and 5 others to reforms in the LabRef database, so we could use the year of legislation in those databases. Of the remaining reform measures, we managed to date 173 records using other – ad hoc – sources.

The number of reforms in the OECD tables is arbitrary for two reasons. First, the tables in the OECD publications contain multiple reforms per cell. Sometimes the distinction between reforms in a single cell is clearly indicated with a paragraph break, often the distinction is less clear. Wherever possible we split up the cells into individual reform measures. A second reason why the number of reforms in the OECD publications is arbitrary is that the OECD publications overlap; one reform is often mentioned in multiple publications. For our regression analysis, this is inconsequential, since we merely need a true/false indicator per country-year combination. The number of reforms in that country-year is irrelevant.

The LABREF database

The LABREF database is managed by the European Commission in cooperation with the Employment Committee (EMCO). The database contains enacted legislation, as well as other public acts of general scope, including measures entailing changes in the implementation framework of a previously adopted measure. In addition, the database encompasses relevant collective agreements and tripartite agreements. The LABREF covers the 28 EU Member States over the period 2000-2013. Hence, only a subset of our 23 OECD countries is included in this LABREF database. We ran regressions with and without the records from the LABREF database, and our results are robust since most reforms included in the LABREF are also in the ISSA and OECD database. From the LABREF dataset we use records from the policy domain “Early Withdrawal”, specifically the policy fields “Disability schemes” and “Early retirement”. From the policy domain “Labour Taxation”, we use the policy fields “Employers’ social security contributions”, “Employees’ social security contributions”, and “Self employed’s social security contributions”. Finally, there is one relevant record on pensions hidden in the “Net replacement rate” of the domain “Unemployment benefits”. Ultimately, the LABREF database contains 88 records that are clear and relevant for our analysis.

Ad-hoc sources

Besides the above four main databases, we consult country-specific publications on the history of pension systems. Wherever possible, we try to link the reforms mentioned in these ad-hoc sources to our four main databases. In the end, we use 114 reforms from these country-specific sources that come on top of the reform measures already extracted from our four main databases.

B. Comparison with other datasets in the literature

Leibrecht and Fong (2017)

Leibrecht and Fond (2017) limit their data to the introduction of second-pillar defined-contribution schemes in a sample of 100 countries over the time period 1980-2012. Of these countries, 31 actually implemented such a system, the other 69 did not. Of the 31 actually-implementing countries, only 3 cases are in our sample (Australia: 1992, Norway: 2002, and Sweden: 2007).

Verbic and Spruk (2019)

Verbic and Spruk (2019) include reforms in 36 countries. Their sample includes all our 23 OECD countries over the period 1970-2013. They include the following data:

- The year(s) in which the first old-age and disability law were introduced
- Year and name/number of the subsequent old-age legislation
- Year and name/number of occupational pension legislation
- Year, name/number, and often a short description, of supplementary pension legislation

The first item is irrelevant for our dataset, since all systems were designed well before 1970, the start of our sample. The other three categories refer to the traditional first, second and third pillar of any pension system. The authors' "primary source of defining, coding and measuring pension reforms is International Social Security Association (http://www.issa.int)", which is also one of the four main databases we also use. For the countries in our sample, our database encompasses the Verbic and Spruk dataset, provided that the reforms affect the government's budget, and classifies the reforms according to two additional dimensions (the effect on the intertemporal government budget constraint, so expanding versus contracting, and the width of the groups affected).

C. Selection of the break in the effect of growth for the "Expanding only" regime

To control for potential structural breaks, we expand the baseline specification by adding a time-specific dummy variable which interacts with the intercept and the business cycle variables $(GROWTH_{it}, DEF_{it}, UNEMPL_{it})'$. We test for a break at the end of each decade, because the identification of a more precise location of a break year would leave an unwarranted impression that there exists a specific year of an OECD-wide change in reform policy. This dummy variable takes a value of one in each year before the break year and a value of zero for the break year and each year after. In only one instance does the coefficient on a business cycle variable jump at a break year, and this is for $GROWTH_{it}$ in the "Expanding only" regime.[30] The next table reports the inverse of the p-value of a Wald test of the baseline specification against a more general one in which we allow for a break in the intercept and in the coefficient of $GROWTH_{it}$ starting with the break year. An inverse p-value of 10 or higher thus indicates significance at the 10% level or lower. The inverse p-value peaks for a break in 1980. Hence, in the sequel for the "Expanding only" regime we impose a break in 1980 and we allow for a different intercept and coefficient on $GROWTH_{it}$ for the years 1970 up to and including 1979. Tests for a second break are always insignificant. The same analysis for "Contracting only" and "Contracting and expanding" never yields significant jumps in the coefficients on the business cycle variables, hence we always estimate the regressions for "Contracting only" and "Contracting and expanding" without a break.

Inverse p-value for Wald test of a break in that year

Break year	Expanding only
1980	38.5
1990	7.7
2000	7.8
2010	14.7

[30] In a preliminary regression, we found a jump in the coefficient of $\Delta OAD25_{it}$. However, this turned out to be the result of two specific years in the sample. Hence, in the following we do not impose a jump on the coefficient of $\Delta OAD25_{it}$.