What Drives Pension Reform Measures in the OECD? Evidence based on a New Comprehensive Dataset and Theory

Beetsma, R.; Romp, W.; van Maurik, R.

Publication date
2017

Document Version
Submitted manuscript

WHAT DRIVES PENSION REFORM MEASURES IN THE OECD?[1]

Evidence based on a New Comprehensive Dataset and Theory

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

Abstract

Using a unique narratively-constructed dataset of pension reform measures of the OECD countries since the 1970s, we explore the determinants of those measures based on information available when they are legislated. We distinguish expansionary, contractionary and combined reform regimes. No regime's likelihood is affected by current or projected demographic changes. By contrast, business cycle indicators play a substantially larger role: a worsening makes contractionary and combination regimes more likely and expansionary regimes less likely. A simple theoretical model with an adjustment cost of changing the pension arrangement can account for reform responsiveness to the business cycle and non-responsiveness to demography.

This version: September 2017

JEL codes: H55, H62, J11, J26.

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

---

[1] We thank seminar participants at the National Bank of Poland and the Dutch Central Bank. This research was made possible with the financial support of 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.

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 the determinants 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 successful pension reform.

This paper makes three contributions to the literature on pension reform. Our first contribution is that, following a "narrative approach" based on reading relevant documents, we construct a new dataset on pension reform measures in OECD countries over the period 1970 – 2013. This is one of the most comprehensive datasets on pension reform measures that exist to date, covering a large set of countries over a long sample period, while it is the first dataset trying to capture all reform measures that affect the generosity and, hence, the cost of pension arrangements. Further, the dataset differs from many 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.

The pension reform measures in our dataset concern legislated measures that in one way or the other affect the public budget. This does not only include changes to public pension provision, but, for example, also measures that stimulate a later take up of a private pension, so that individuals are employed and pay taxes for longer. We categorize the reforms into "expansionary" and "contractionary" and define three regimes. The "Expanding only" regime is characterized by measures that increase coverage, eligibility and/or the pension benefit level, while the "Contracting only" regime is characterized by measures aimed at increasing financial and fiscal sustainability and/or stimulating work incentives. Finally, there is the "Expanding and contracting regime", which prevails when expansionary and contractionary measures occur in the same year. We find that over time expansionary reform activity becomes less frequent, while the incidence of contractionary measures and expansionary and contractionary measures happening simultaneously increases.

Our second contribution is that, using our new dataset, we explore the determinants of pension reform measures in a systematic econometric analysis that links the reform regimes to the demographic, economic and budgetary information available at the moment of their legislation. We also investigate the role of political variables, economic and financial crises and the presence of supranational fiscal constraints for reform. Econometric analyses on this scale of the pension reform determinants are sparse, if they exist at all. None of the three regimes are affected by current or projected future demographic changes. This is remarkable, as we would a priori expect reform measures to be closely linked to long-run financial sustainability considerations. By contrast, we find that business cycle indicators – broadly defined, so capturing economic growth, unemployment and the state of the public finances – play a substantially larger role. In particular, a worsening of the business cycle enhances the likelihood of the "Contracting only" and the "Expanding and contracting" regimes, while it reduces that of the "Expanding only" regime during the second part of the sample period. Anecdotal evidence supports these 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. A year earlier, 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. There is also anecdotal evidence of favorable circumstances being conducive to expansionary reform. An example is Belgium in 1997, when real GDP growth was expected to accelerate to $2^1/_4$% (OECD, 1997), with the introduction of a minimum pension amount for each year in employment for at least one-third of the normal working time.

As our third contribution, we construct an original theoretical framework that can account for the observed empirical regularities. The model features a lump-sum adjustment cost of the changing an existing pension arrangement. Numerical analysis shows that the model can replicate the responsiveness of pension arrangements to the business cycle and their non-responsiveness to current and projected demographic developments.

Our paper connects in different ways to the literature. First, while systematic econometric analysis of pension reform measures is rare, recently some papers constructed datasets of pension reforms. In particular, after constructing our own dataset we became aware of Spruk and Verbic (undated) and Leibrecht and Fong (2017), who construct similar, but less detailed datasets.[5] The focus in the empirical analysis of Spruk and Verbic is on the political determinants of reforms and in Leibrecht and Fong on the political, economic and social determinants of retirement income privatization. In contrast to their results, we find that political and demographic variables turn out to essentially play no role for pension reform measures, while variables indicating the state of the economy are all important. The fundamental difference is that, in contrast to these two papers, we link reform activity to *changes* in (projected) old-age dependency ratios and *not* their level. Changes

[5] Spruk and Verbic (undated) construct a dataset for 34 countries over the period 1970-2013, focusing on the transition from unfunded to funded pension schemes. Leibrecht and Fong (2017) for a broader cross-section of countries, but a shorter time period, identify the privatization of retirement income systems. Crucial differences with our dataset are that (1) we try to capture any reform measure that affects the generosity of pension arrangements, leading to a substantial increase in the number of observations, and (2) we classify these measures into whether they make the arrangement more or less generous.

are crucial, because only a change can explain why a reform occurs now and not at some arbitrary other moment. A final difference is that we also provide a theoretical explanation for our finding that pension reform measures are mostly explained by the state of the economy, and not by (projected) demographic changes.

Our paper relates also to other work exploring the determinants of pension arrangements and the drivers 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 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. 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*. (2003) 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 to reform 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 a deepening of the capital markets, increased private savings and higher economic growth.

Our main empirical finding – the timing of pension reform measures is related to the business cycle – is closely related to the crisis-induced reform literature (Rodrik, 1996, Abiad and Mody, 2005, Bonfiglioli and Gancia, 2015, Ranciere and Tornell, 2015, Mahmalat and Curran, 2017). Thompson

[6] Other suggested factors than those discussed are 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).

(2009) finds that structural reforms are typically legislated during periods of poor economic performance. However, our analysis differs in fundamental ways from that in the crisis-induced literature. While the latter 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 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 reform measures and expansionary reform measures. We find that the timing of contractionary reform measures is related to a weak economy. Expansionary reform measures, which include such structural reform measures as an expansion of the coverage of women, tend to be implemented during economic upswings. Such expansionary reform measures are not considered in the crisis-induced literature. We rationalize the timing of both contracting and expanding reform measures in a model with implementation costs, where the net gains of reform measures fluctuate with the economic situation, not the political constraints that prevent reforms.[7]

The remainder of the paper is organized as follow. Section 2 presents the data. Section 3 sets up the empirical framework, while Section 4 reports and discusses the estimation results. In Section 5 we construct and numerically analyze our theoretical framework to account for our empirical findings. Further technical details, details on the construction of our data and the results of a number of robustness tests are available from our homepages.

2 The data

2.1 The data on pension reform measures

Our 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 documents from the International Social Security Association (ISSA, 2014), the OECD (2012, 2013) and the International Labor Organization (ILO, 2014), both of which provide information on (legislation on)

[7] There is a literature on political and legal constraints that prevent policymakers to pursue 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 crisis, while bank regulators were able to force banks into recapitalizing themselves after the start of the crisis.

social security reforms. In cases where additional information is needed, we consult other, mostly national sources. A full list of all the consulted sources is attached to the dataset. Reforms comprise both smaller (parametric) and more fundamental changes to pension arrangements. We do not try to make an explicit distinction between small and large reform measures, to avoid the danger of being subjective in making such a distinction.

We date reform measures 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 was made. It is obviously conceivable that in many instances the discussion about the reform started in the year before the reform was legislated, or possibly even earlier. However, it is practically unfeasible to uncover for each reform the moment when such a discussion was started, while deviating from the year of legislation in recording reform measures would introduce another source of arbitrariness. In our empirical analysis, we will show that our findings are robust to using moving averages of the baseline variables used to explain reform measures. Hence, this issue seems to be of only limited relevance.

We divide reform measures that are relevant for our analysis into four categories: (1) "Coverage", the number of reform measures that expand the coverage of the pension arrangement, for example by loosening the eligibility criteria; (2) "Generosity and adequacy", the number of reform measures that expand the generosity of the pension system, for example by raising the benefit level; (3) "Financial and fiscal sustainability", the number of reform measures that enhance the financial sustainability of the pension arrangement, for example by reducing benefits or by raising the retirement age; (4) "Work incentives", the number of reform measures that enhance work incentives, for example by introducing bonuses for working after the minimum age at which pension benefits can be collected. Besides these reform categories, there are various other possible measures, for example pertaining to financial safety, that do not have a direct or clear effect on the public finances. These are not interesting for our purposes and, hence, we do not separately categorize them. Examples of how we classify measures on the basis of available text passages are found in Appendix A.

The categories "Coverage" and "Generosity and adequacy" can be considered as expansionary of the pension system, because, if a reform in either of these two categories takes place, (long-run) pension obligations will increase. By contrast, the categories "Financial and 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 public budget. Reading the relevant texts it is not always unambiguously clear how to classify the reform measures. However, the fact that we always merge the observations of "Coverage" and "Generosity and adequacy” into a single category and those of “Financial and fiscal sustainability” and “Work incentives” into a single category limits the consequences of potential judgment errors in the classification of reform measures.

Aggregating across the sample countries, Figure 1 depicts the number of each type of reform measure by sample year. To clarify the figure, take as an example the year 1970. In this year a total of four reform measures have been legislated in our set of sample countries, three in the category “Coverage” and one in “Financial and fiscal sustainability”. We observe some tendency towards an increasing number of reforms as time progresses. Early in the sample period, the reform measures are predominantly of the expansionary type, while in later periods they tend to be mostly of the contractionary type.

We create two dummy variables. The dummy “Expansion” is one if at least one reform measure falls within the categories “Coverage” or “Generosity and adequacy”, and zero otherwise. The variable “Contraction” is one if at least one measure falls within the categories “Financial and fiscal sustainability” or “Work incentives”, and zero otherwise. Using our 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 “Expanding and contracting”, 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 “Expanding and contracting”, 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. This interpretation is supported by the fact that in the majority of the cases when expanding and contracting 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 (2014) documentation.

Table 1 lists for each sample country the number of reform measures in each category. There are substantial differences in the numbers of measures carried out by our sample countries. France and Greece, with 54 and 44 measures, respectively, enacted the largest number, while Iceland and Norway, with 7 and 11 measures, respectively, enacted the smallest number.

Table 2 reports the number of reform measures in each 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 – 1991 and 1992 – 2013). The total number of expansionary reform measures is 110+166=276 distributed over 244 (country, year) - combinations, implying that for some countries in some years there is more than one expansionary measure, while the number of (country, year) -

combinations with expansionary reforms *only* is 184. In other words, there are 244-184=60 (country, year) - combinations that fall into the “Expanding and contracting” regime. Further, the total number of contractionary reforms is 133+63=196, distributed over 159 (country, year) - combinations, while the number of (country, year) - combinations with contractions only is 99. This confirms that there are 159-99=60 (country, year) - combinations that fall into the “Expanding and contracting” regime.

We observe that there is a reasonable balance of expansionary measures over the two sub-periods, although the second sub-period features slightly more of them. By contrast, contractionary measures are far more prevalent in the second sub-sample than in the first sub-sample. Only a quarter of the measures categorized as “Financial and fiscal sustainability” and only a tenth of the measures categorized as “Work incentives” take place in the first sub-period. Further, about three-quarters of the instances of the “Contracting only” regime occur in the second sub-period, while the same holds for the regime of “Expanding and contracting”.

Table 3 provides further details on the “Expanding and contracting” regime, where we report the numbers of the various combinations of expansionary and contractionary measures taking place in the same year. The total number of 94 combinations exceeds the number of 60 (country, year) – pairs in which the “Expanding and contracting” regime is observed. The reason is that there are instances of packages that contain more than one combination. For example, Australia in 2013 enacted a package of six reforms covering all four possible combinations, thus adding one more observation to each of the four combinations. The frequencies of the combinations are quite well in line with the relative overall frequencies of the two types of expansionary measures and the two types of contractionary measures. The highest frequency is formed by the combination of “Generosity and adequacy” and “Financial and fiscal sustainability”, which is in line with the fact that “Generosity and adequacy” is more prevalent than “Coverage” and that “Financial and fiscal sustainability” is more prevalent than “Work incentives”. Hence, compared to what is “normal” Table 3 does not reveal an obvious combination of contractionary and expansionary measures that is preferred by the policymakers.

2.2 The demographic variables

We use two demographic variables in our analysis: the change of the current old-age dependency ratio and the change of the projected 25-year ahead old-age dependency ratio. 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. Each issue reports the current value and projects the future old-age dependency ratio, with projections done for intervals of 5, 10 or 15 years. The projections furthest into the future range from twenty-two years ahead to a century ahead, depending on the issue. For the missing years we interpolate the ratio using the surrounding years for which we do have projections.

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 European Commission's Ameco dataset, the IMF World Economic Outlook and the World Bank.

2.4 The political variables

Our political variables are obtained from the Comparative Political Data Set I (Armingeon *et al.*, 2015). 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 crisis indicators taken from Laeven and Valencia (2012) and dummy variables to indicate the participation (or not) in the European Union (EU) as of 1992,[8] which is the year when the Maastricht Treaty was signed, or the Eurozone as of the year of entry.

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. We will try to explain reform decisions on the basis of real-time information available at the moment when reform measures are enacted. Our baseline specification will be a logit regression that links the occurrence of a reform regime to the projected change in the old-age dependency ratio, 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.[9] We thus consider GDP growth, the public deficit and unemployment all as indicators of the current state of the economy. We include these

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

[9] Elaborate prior experimentation shows that these are the (only) variables for which there is a systematic role in explaining our reform regimes.

variables jointly in our regression, because their correlation is far from perfect. For example, labour 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.

Our empirical approach deviates in a potentially important way from that suggested by the 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,[10] while in our 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. Later, we will also estimate generalizations of our baseline specification in which we include additional variables to explore the relevance of potential other driving forces behind the reform regimes.

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

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

where $z_{it,r}$ is determined by a reform-regime specific function f_r of the explanatory variables:

$$z_{it,r} = f_r(BASEVAR_{it}, ADD_{it}), \tag{2}$$

with

$$BASEVAR_{it} = (\Delta OAD_{it}, \Delta OAD25_{it}, GROWTH_{it}, DEF_{it}, UNEMPL_{it})' \tag{3}$$

the vector of baseline explanatory variables and ADD_{it} a vector of additional variables. Here, ΔOAD_{it} is the change in the current old-age dependency ratio between the years *t*-1 and *t*, $\Delta OAD25_{it}$ is the change in the projected 25-years ahead old-age dependency ratio between the years *t*-1 and *t*, $GROWTH_{it}$ 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. All explanatory variables will be measured in per cent or in percentage points, such as in the case of ΔOAD_{it} and $\Delta OAD25_{it}$. For each possible reform regime r, we will run a separate logistic estimation, so that the alternative to ending up in regime r is to end up

[10] 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.

in a regime in which no reform measures are taken. The motivation for this approach is that for the various regimes we identify some difference in the years for which we detect structural changes in the relationship between the reform measures and the explanatory variables. 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.

In the following, we will always assume that the relationship between $z_{it,r}$ and our set of explanatory variables is (piecewise) linear. We will also always include country-fixed effects.[11] However, we do not include of time-fixed effects. They would lead to rather uninformative results, as the explanatory variables tend to be rather highly correlated across the OECD countries.

4 Results

This section describes and interprets the results of our regression analysis. However, before turning to the estimations, we will test for the existence of breaks in the relationship between the explanatory variables and the incidence of the various regimes.

4.1 Break tests

The breakpoint testing procedure involves estimating a logit regression for regime *r* in which we specify

$$z_{it,r} = \alpha_{0i,r} + D_y\gamma_{0,r} + \left(\alpha_r + D_y\gamma_r\right)' BASEVAR_{it} + \delta_r' ADD_{it}, \qquad (2')$$

where D_y is a dummy that takes a value of zero before year y, and of one as of year y, the dummy's coefficient is $\gamma_{0,r}$, $\alpha_{0i,r}$ capture country-fixed effects, and α_r and γ_r are coefficient vectors of the appropriate dimensions. Hence, we allow for a break in the intercept as well as the coefficients of the explanatory variables as of year y. To avoid the number of coefficients from becoming too large, we do not allow the coefficients on the control variables to shift between the sub-sample periods. We let y run over the years 1980 – 2003, implying that we require a minimum of 10 years before and after a potential break. For each of our three reform regimes, Figure 2 depicts the reciprocal of the *p*-value associated with the hypothesis that $\gamma_r = 0$, while varying *y* over the years. We denote the year in which the test statistic reaches its minimum p-value by y^*. We find that $y^* = 1988$ for the

[11] Including country-fixed effects in a panel logit regression produces biased coefficients (Chamberlain, 1980). To avoid this bias, we also estimate the coefficients of our explanatory variables through conditional logit. Comparing the estimated coefficients of the explanatory variables under conditional and regular panel logit shows that these are virtually identical, indicating at most only a very small bias. The drawback of conditional logit is that this method does not generate the potentially large country-fixed effects needed for the mean marginal effects.

"Expanding only" regime, $y^* = 1992$ for the "Contracting only" regime and $y^* = 1997$ for the "Expanding and contracting" regime. Figure 2 shows that the break points can be very clearly identified for each of the regimes.

4.2 Baseline estimates

Our baseline specification consists of equations (1) – (3), with D_y set to D_{y^*}, therefore allowing for a change in the coefficients of the explanatory variables in the year for which we identified a break above. Hence, each reform regime features a different break year. In our baseline, we try to explain the reform regime on the basis of current and projected changes in the old-age dependency ratio, GDP growth in deviation from its national average over time, the deficit and the unemployment rate.

Table 4 presents the coefficient estimates for our baseline specification for the "Expanding only" regime. Column (1) shows that in the period before 1988 none of the estimated coefficients are significant.[12] The period as of 1988 reveals some differences compared to the period before. First, the coefficient on the change in the projected future old-age dependency ratio exhibits a statistically significant fall and, hence, becomes negative. However, the Wald test that this coefficient is zero in the second sub-period cannot be rejected, suggesting that a projected increase in future ageing fails to exert a significantly negative effect on expansionary reform. Second, the coefficient associated with the GDP growth variable exhibits a significant upward jump and becomes significantly positive at the 1% level in the second sub-period, indicating that policymakers now also pay attention to the state of the economy when considering whether or not to expand pension arrangements: higher-GDP growth (relative to average growth over time) leads to more expansionary pension measures. This link may be driven by the fact that higher growth generates higher revenues from pension contributions or tax payments, thereby providing room for expansion. 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 estimates presented in Table 4 do not provide us with direct information on the magnitude of the effect of a marginal increase in an explanatory variable on the likelihood that the "Expanding only" regime will materialize. As can be seen from model (1) – (3), the size of the effect depends on the values of the explanatory variables at which it is evaluated. Table 5 reports for the "Expanding only" regime the mean marginal effects of changes in the explanatory variables calculated at their averages over the full sample or the relevant subsample. We observe that for the sub-period after 1988 a one-percentage point increase in the growth rate of the economy (relative to its average over time) raises the likelihood of the "Expanding only" regime by 2.1 percentage points.

[12] Here, and in the sequel, we mean by significance that a coefficient is significant at the 10% or a higher level.

While the effect of a change in this variable during the second sub-period is significant, its magnitude is rather limited.

The estimates for the "Contractions only" regime also fail to provide any evidence of a role for the change in the current and projected old-age dependency ratio, neither in the first nor in the second sub-sample. This may be a bit surprising as many countries have started raising official retirement ages, allegedly motivated by rising current and future life expectancy. Yet, we do observe an important role for the state of the economy, as captured by the growth variable, which in both sub-periods exerts a significantly negative effect (at the 5% level) on the likelihood of a "Contracting only" regime. As Table 5 reports, the mean marginal effects during the second sub-period, again evaluated at the sub-sample averages of the independent variables, on the likelihood of the "Contracting only" regime are -1.6 percentage points for a one-percentage point increase in the economy's growth rate and 1.5 percentage points for a one-percentage point increase in the unemployment rate. We also observe that, going from the first to the second sub-period, the size of the mean marginal effects roughly triples in absolute magnitude.

Our final set of baseline regressions concern the "Expanding and contracting" regime. The coefficient estimates are reported in Column (3) of Table 4. The constant exhibits an upward jump in 1997. Again, the current and projected changes in the old-age dependency ratio play no role in explaining the likelihood that an "Expanding and contracting" regime materializes. As before, only the current state of the economy plays a role, this time captured by the effect of the public deficit. One may ask whether pension reform measures have potential feedback effects on the deficit, thereby biasing its coefficient. However, the effects of reform measures are unlikely to have a material feedback effect, as they are likely dominated by all the other changes in public spending, while the effects of contractionary measures tend to be mostly felt in the longer run. Table 5 shows that the mean marginal effect of a one-percentage point increase in the public deficit raises the likelihood of the "Expanding and contracting" regime by 0.6 percentage points in the first sub-period and by 1.0 percentage points in the second sub-period.

4.3 Robustness[13]

4.3.1 Multinomial logit

The above regressions ignore the fact that our reform regimes, including the "No reform" regime in which no reform measure is undertaken, are mutually exclusive. This fact can be exploited using a

[13] For brevity, we do not include all regression results in this subsection; where necessary, we mention the significant coefficients. Quantitatively, the change in the coefficients of the baseline variables is always negligible. All regression results are available upon request.

multinomial logit model. However, the drawback is that, if we want to allow for a structural break, this break has to occur at the same date for all possible regimes, while the above regressions based on the logit model clearly indicated different breakpoints for the various regimes. For this reason, we employed the standard logit model for our baseline. Here, we check the robustness of our baseline results by repeatedly estimating the multinomial logit model, while varying the common breakpoint for our three reform regimes.[14] Table 6 reports the estimates. Columns (1), (2) and (3) report the coefficient estimates for “Expanding only”, “Contracting only” and “Expanding and contracting”, assuming that the common breakpoint for the three reform regimes is 1988, 1992 and 1997, respectively. Hence, these breakpoints correspond to those found earlier for our three reform regimes. In each instance we only report the coefficient estimates for the reform regime we are interested in. We observe that the estimates are very similar to those from the ordinary logit estimations. Using the information from the Wald tests, the coefficients and standard errors of the variables that were statistically significant before remain significant and hardly change in magnitude. Moreover, there are no instances of variables that are significant under multi-nominal logit, but not significant under the ordinary logit estimation.

4.3.2 Averages of explanatory variables

The analysis so far has implicitly assumed that the recorded reform measures are driven by state of the economy and the demographic factors in the year they are legislated. However, the design and the legislative process underlying many reform measures, especially the larger ones, may take some time to materialize. Hence, on the one hand the relevant information set that forms the input for a reform decision may also include realizations of relevant variables in earlier years. To deal with this possibility, we allow our reform regimes to be explained by averages of current and past values of our explanatory variables. On the other hand, including too many lags into these averages would render it difficult to detect any effects on the likelihood of reform measures. Hence, we choose two lags as a reasonable compromise. Our specification of $z_{it,r}$ in (1) now becomes:

$$z_{it,r} = \left(\alpha_{0i,r} + D_y \gamma_{0,r}\right) + \tfrac{1}{3}\left(\alpha_r + D_y \gamma_r\right)' \sum_{j=0}^{2} BASEVAR_{i,t-j}.$$

Table 7 repeats the baseline regressions reported in Table 4. We observe that the projected old-age dependency ratio is significant for both “Expanding only” and “Contracting only” in the first sub-sample. The signs of the coefficient, positive and negative, respectively, may be not be as expected,

[14] 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 reform regimes “Expanding only”, “Contracting only” and “Expanding and contracting”. Hence, $R = 3$. The likelihood of ending up in the “No reform” regime is $1 - \sum_{h=1}^{R} p_{it,h}$.

but a potential reason for these signs may be that projected increases in life expectancy were felt to necessitate the creation of better pension provisions for a growing group of future elderly. However, as shown by the Wald test statistics, the projected old-age dependency ratio loses in both cases significance in the second sub-period. The deficit is significant with a negative sign for the “Expanding only” regime in the first sub-period, but it loses significance in the second sub-period. Also, in contrast to the baseline regression, for this regime GDP growth is insignificant in the second sub-period. For the “Contracting only” regime, the second-period GDP growth rate is again highly significant, while unemployment remains significant at the 10% level, indicating that the state of the economy continues to play an important role in determining the likelihood of reform measures. Similarly, for the “Expanding and contracting” regime, the effect of the deficit on the likelihood of reform continues to be significantly positive.

4.3.3 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 of trade ($OPENNESS_{it}$), CPI inflation ($INFLATION_{it}$), general government debt as a share of GDP ($DEBT_{it}$) and the average of the short-term (3-month) and the long-term (10-year) public debt yield ($INTEREST_{it}$). For the sake of space, we do not report the estimates. The only coefficient that is significant is that of $OPENNESS_{it}$ for “Expanding only”: an increase in openness makes expansionary measures less likely, possibly because in more open economies it is more important to restrain labor costs in order to remain competitive. For the other two regimes, none of the additional controls are significant. Baseline variables that were significant (insignificant) before remain so. Hence, we conclude that the state of the economy remains the main driving force behind pension reform measures.

4.3.4 Supranational budgetary constraints

Supranational budgetary constraints could motivate the adoption of reform measures (Bertola and Boeri, 2002), especially those reforms 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 EU countries that have tried to meet the criteria for accession to the Eurozone in the run-up to Economic and Monetary Union during most of the nineties and that have been bound by these criteria since they entered of the Eurozone. These criteria constitute an upper-bound of 3% on the public deficit – GDP ratio and an upper-bound of 60% on the public debt – GDP ratio (e.g., see Beetsma and Uhlig, 1999, Beetsma and Debrun, 2007). 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. To investigate their influence on the likelihood of reform measures, we construct a dummy $D\_MAASTRICHT_{it}$ that takes on a value of one for each EU country as of 1992, i.e. after the Maastricht Treaty was signed, if it was already an EU member in that year, or for countries that joined the EU later as of the year of joining. The dummy is zero in all other cases. Similarly, we introduce a dummy $D\_EUROZONE_{it}$ that is one for each Eurozone member as of the year it joined the Eurozone and zero, otherwise.

Again, for the sake of space, we do not report the estimates. The Maastricht dummy is never significant. The Eurozone dummy is significantly positive for the "Contracting only" regime, suggesting that the fiscal constraints under the Eurozone may have played a role forcing governments to adopt contractionary measures. The GDP growth rate remains significant in both sub-periods, although its significance weakens a bit in the second sub-period. Hence, the conclusion that the current state of the economy is an important driving force behind pension reform remains unaltered.

4.3.5 The role of political variables

Spruk and Verbic (undated) point to the potential role of the ideological leaning of the government and the possible role of elections.[15] This subsection tries to address the role of these and other political factors. We do this by each time expanding the baseline regression by one political variable and checking whether the coefficient estimates under the baseline have undergone material changes and whether the political variable is significant. The addition of the political variables does not affect the coefficient estimates of the baseline explanatory variables. The Wald tests for the significance of their second-period values are always the same as under the baseline. Overall, the evidence of the state of the economy as a major main driver of pension reform measures remains strong.

The political variables are significant in only one case. An increase in RAE_LEG_{it}, the degree of fractionalization of the parliament, exerts a positive effect on the chances to expand the pension system. This may be explained by the possibility that with more fractionalization there is a larger need for bargains in which the various fractions in parliament get something to their liking. In all other instances, the political variables play no role. This is the case, in particular for (1) $\mathrm{CABINET\_INDEX_{it}}$, 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 left- to right-wing parties; (2) $\mathrm{GOV\_PARTY_{it}}$, which captures the political color of the government: the higher it is, the stronger is the position of the left-wing parties in government; (3) GOV_GAP_{it}, which captures the ideological difference between the new and old cabinet: a higher value implies a larger left-ward ideological shift of the government; (4) $PARLIAMENT\_INDEX_{it}$, which captures the average political color of the parties in parliament: a

[15] 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.

higher value implies a more right-wing average orientation of the parliament; (5) $D\_ELECTION_{it}$, a dummy for an election year; (6) $\mathrm{GOV\_NEW_{it}}$, a dummy for whether the government is new; (7) $\mathrm{GOV\_CHAN_{it}}$, the number of government changes in a year; and (8) GOV_TYPE_{it}, which indicates the strength of the government party/parties in parliament.[16]

4.3.6 Controlling for crises

In our final 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 (2012) we define a dummy D_CRISIS_{it} that takes a value of one in all the cases in which they identify the occurrence of either a banking crisis, a currency crisis or a sovereign debt crisis, as well as in the case a sovereign debt restructuring takes place. Because their dataset only spans the period up to 2011, our sample period is now ends in the year 2011. The number of times a crisis coincides with one of the reform regimes is rather limited.[17] Hence, one should be careful not to over-interpret any of the results in this subsection. The crisis dummy is only significant (at the 10% level) for the "Contracting only" regime, in which case it exerts a positive effect on the likelihood of a contractionary reform. In this case, the significance of the growth rate in the second sub-period weakens, likely because crisis periods capture a substantial fraction of the periods in which growth is low. In none of the other cases anything changes to the significance of the baseline explanatory variables, indicating that the state of the economy (including the potential situation of a crisis) remains a major driving force behind pension reform measures.

5 A theoretical framework

In our empirical analysis, we found at most only very limited evidence of projected future changes in the old-age dependency ratio on pension reform measures. By contrast, there was rather substantial evidence of the current state of the economy, broadly defined, on reform measures. This section presents a simple theoretical framework that can simultaneously explain why the cyclical state of the economy can trigger adjustments in pension generosity, while projected changes in the old-age dependency ratio have less power in doing so.

---

[16] In view of the possibility that reforms are triggered only when the circumstances change, we have also explored whether the results are affected by including (one at a time) the changes in the political variables (except for those in GOV_GAP_{it} and GOV_NEW_{it}, which are already differences, and $D\_ELECTION_{it}$. Only the change in RAE_LEG_{it} is significant (at the 10% level) in the "Expanding only" regression and the change in GOV_TYPE_{it} (at the 10% level) in the "Expanding and contracting" regression. Again, there are no changes in the coefficient estimates of the baseline explanatory variables, and the outcomes of the Wald tests for their significance in the second sub-period are unchanged.

[17] The exact numbers of coincidences are 18 for crisis and "Expanding only", 21 for crisis and "Contracting only", and 13 for crisis and "Expanding and contracting".

There is a political party that runs the current government and discounts the future at a factor $0 < \pi < 1$. This factor captures both the government's innate time preference as well as a potential reduction in its effective discount factor resulting from the possibility of losing office to a political competitor. The current government cares – among other things – about the income position of the elderly as measured by the pension pay-out $P > 0$.[18] The latter is chosen taking into account current and future economic and demographic conditions. Current economic conditions are fully summarized by an exogenous stochastic business cycle indicator $Y > 0$. It follows a Markov process with a stationary distribution. Hence, all the shocks to Y are of a temporary nature.

The demography in our theoretical framework is captured by the old-age dependency ratio, which is defined as the ratio of all retirees over all workers. Ideally, forecasts about the old-age dependency ratio would be based on the population pyramid and the fundamental demographic forces: fertility, mortality and migration. In most countries, the current size of each cohort is well known. Levels of fertility and mortality change only slowly over time. Thus, when fertility or mortality is high in one year, this is likely also the case the next year. This implies a strong, but not perfect, serial correlation in these two components. International migration is more volatile, but some degree of serial correlation should be expected, because the driving economic, legal, political and social conditions tend to change only slowly over time (e.g., Preston *et al.*, 2000). In addition, in most developed countries, migrant flows in the various age categories are small relative to the existing cohorts. All these factors render the old-age dependency ratio relatively predictable in the short to medium term.

Modelling the three fundamental sources of demographic change explicitly and keeping track of all the cohorts is well beyond the scope of this paper. The curse of dimensionality would cause the state space to explode. Instead, we try to capture the current and projected future demographic situation with only the current old-age dependency ratio B and the long-run level b to which B is expected to converge. Such a long-run level exists if the demography is stable, which is the case when migration flows, age-specific fertility and age-specific mortality rates are constant. In view of the continuing medical progress, especially this last assumption seems unrealistic. However, a constant link of the retirement age to life expectancy would continue to produce a constant long-run age-dependency ratio even when mortality rates are falling. We allow for temporary fluctuations in migration flows, fertility and mortality reduction by assuming that the pair (b, B) follows a (highly-persistent) time-stationary Markov process. We also assume that *b* and *B* are independent of *Y*.

[18] The positive value of P implies that transfers from the old to the young are excluded.

We assume that the government's instantaneous utility is fully determined by current economic and demographic conditions. Hence, it can be written as $U(P, Y, B)$. We exclude financing of the pension system with government debt, so there exists an upper-bound $\bar{P}(Y, B) < \infty$ on P, which is determined by Y and B. Instantaneous utility is strictly concave in P, and increasing and concave in Y. Strict concavity in P arises, for example, in a situation in which an endowment needs to be divided over various generations (a higher pension payout is at the cost of working generations) or when a higher pension payout reduces resources available for other public spending. For simplicity, we assume that U is continuous and twice differentiable in all its arguments. We denote the pension benefit that maximizes instantaneous utility by

$$\tilde{P}(Y, B) = \underset{P}{\operatorname{argmax}}\, U(P, Y, B).$$

The literature (see, e.g., Gonzalez and Eiras, 2008, and Ciurila and Romp, 2015) suggests that there are two opposing forces of the old-age dependency ratio B on the optimal pension payout chosen by a politician. A higher old-age dependency ratio raises the cost of the pension system, putting downward pressure on the individual pension payout, but it also increases the electoral weight of the group of retirees, which causes upward pressure on the individual payout. Ciurila and Romp (2015) conclude that in a probabilistic voting setting an office-seeking politician will divide the financial burden of a higher old-age dependency ratio over all generations by increasing the contribution and at the same time cutting the *per-retiree* pension benefit. This suggests a downward sloping relationship between B and the optimal individual pension pay-out $\tilde{P}$, so $\tilde{P}_B(Y, B) \leq 0$, which requires $U_{PB} \leq 0$. Finally, we assume that windfall gains are divided over both the active and retired part of the population, hence $U_{PY} \geq 0$, so $\tilde{P}_Y(Y, B) \geq 0$. A special case that satisfies these assumptions is when the government maximizes $U = f(G) + Bv(P)$ subject to $G + BP = Y$, where the functions $f(\cdot)$ and $v(\cdot)$ are continuous, increasing and strictly concave, G is an amount of a public good and Y is an exogenous endowment. Hence, the government aims at optimally allocating a given endowment over public good and pension provision. Another special case, which we will use later, is where $\tilde{P}$ is proportional to Y and instantaneous utility only depends on the optimal payout relative to the current payout, so instantaneous utility can be written as $U = u\big(Y\tilde{P}(1, B)/P\big)$, with $u(\cdot)$ strictly concave and $\tilde{P}(1, B)$ positive, but decreasing in B. For now, we continue with the more general setup.

Changing the pension benefit P comes at a fixed utility cost $K > 0$ to the government, which is the same irrespective of whether the benefit is raised or reduced. There is no cost associated with keeping the existing benefit unchanged. Note that this specification rules out possible electoral costs of changing the benefit through a lower probability of being re-elected. We consider a discrete time setting, hence the government is free to implement a change at fixed moments, as long as it pays the fixed cost K. The government's optimization problem is described by

$$V(P,Y,B,b) = \max_{P'}\ U(P',Y,B) - I(P'-P)K + E[V(P',Y',B',b')|Y,B,b] \tag{4}$$

where $I(\cdot)$ is an indicator function, such that $I(x)=0$, if $x=0$, and $I(x)=1$, if $x\neq 0$. The expectation of the continuation value is well-defined due to the assumptions that Y and (B,b) follow Markov processes. The government's problem of selecting the optimal pension benefit P' is comparable to a standard optimal pricing problem with menu costs or any other (s,S)-type of model, but with one additional complexity: the government has additional information concerning the next period's situation via the long-term old-age dependency ratio b.

5.1 Necessary and sufficient conditions for change and no change of benefit

Each period the government compares the value of changing the pension system now (V^N) with the value of postponing a change (V^P). The government sets a new level $P'\neq P$ for the pension benefit if and only if $V^N > V^P$. We use the convention that in the case of equality, the pension benefit remains unchanged. The value of changing the system now is

$$V^N(Y,B,b) = \max_{P}\ U(P,Y,B) - K + \pi E[V(P,Y',B',b')|Y,B,b], \tag{5}$$

with $V(\cdot,\cdot,\cdot,\cdot)$ as defined in (4). The value of retaining the current pension benefit includes the costs K of future adjustments, and is given by

$$V^P(P,Y,B,b) =\ U(P,Y,B) + \pi E[V(P,Y',B',b')|Y,B,b]. \tag{6}$$

Using the maximum instantaneous gain defined as

$$F(P,Y,B) \equiv \widetilde{U}(Y,B) - U(P,Y,B) \geq 0, \tag{7}$$

we can formulate sufficient and necessary conditions for a change.

Proposition 1: If $F(P,Y,B) \leq (1-\pi)K$, then $V^P(P,Y,B,b) \geq V^N(Y,B,b)$, so postponing is optimal. Hence, $P' = P$. If $F(P,Y,B) > (1+\pi)K$, then $V^N(P,B,b) > V^P(P,Y,B,b)$, so implementing a change is optimal. Hence, $P'\neq P$.
Proof. See Appendix B.

Note that both sufficient conditions only depend on the *current* business cycle situation (Y), the *current* old-age dependency ratio (B) and the *current* pension payout (P). The demographic forecast b is potentially only relevant in the region not covered by the two inequalities in the above proof, i.e. the region for which $(1-\pi)K < F(P,Y,B) \leq (1+\pi)K$.

The intuition behind Proposition 1 is the following. Consider a government who inherits a pension benefit that is "currently fairly optimal" given the current state of the business cycle and the current old-age dependency ratio. That is, the maximum instantaneous gain does not exceed $(1-\pi)K$. At the same time this government is aware of the fact that the demographic forecasts are such that the inherited pension benefit is unsustainable in the future. Hence, it faces two options: change the pension benefit now to make the pension system future proof or keep the pension benefit at the inherited level and change it in the future. Postponing the change in pension benefit has three advantages over the first option. First, the government also postpones paying the fixed cost. Second, the government does not have to set a pension benefit that potentially lowers the current instantaneous utility. Third, in the next period, the government can freely choose the pension benefit that is optimal from that period onwards. Clearly, provided that the inherited benefit is "fairly optimal", it will never be optimal to change the pension benefit now in order to ensure sustainability of the system in the future. In other words, information about the future is irrelevant in this case.

A similar argument holds if the current economic situation is such that the inherited pension benefit is "currently far from optimal". That is, the maximum instantaneous gain exceeds $(1+\pi)K$. In this case, the government will always want to change the pension benefit, irrespective of the demographic forecasts. In taking a decision about the new benefit level, it may take knowledge about the future into account. On the one hand, if both the current state of the business cycle and the demographic forecast are such that the pension benefit should be cut, the government may want to set a new level that reduces the chance of another cut in the future, so as to avoid paying the adjustment cost again. On the other hand, if the current situation asks for an unsustainable increase in the generosity of the pension benefit, the government knows that it will have to change the pension benefit again in the next period. Hence, it might just as well choose $\tilde{P}$.

5.2 Implementation

To implement our model, we make five assumptions:

1. Instantaneous utility of the government only depends on the ratio of the optimal pension benefit $\tilde{P}$ and the current level of the pension benefit P.
2. This optimal benefit $\tilde{P}$ is proportional to the business cycle indicator Y.
3. The logarithm of the business cycle indicator follows an AR(1) process:

$$\log(Y') = \phi_y \log(Y) + \varepsilon, \quad \varepsilon \sim N(0, \sigma_\varepsilon^2).$$

4. The current old-age dependency ratio gravitates geometrically to its long-run value:

$$B' = \lambda B + (1-\lambda)b, \quad 0 < \lambda < 1.$$

5. The long-run old-age dependency ratio takes on discrete values $b^i \in (0,1)$ and features constant transition probabilities p_b^{ij}:

$$Prob\left(b' = b^j \middle| b = b^i\right) = p_b^{ij}.$$

The strongest assumption is Assumption 4. This clearly violates the non-monotonicity over time in the forecasts of the old-age dependency ratio. In most countries, the old age-dependency ratio is expected to peak between 2030 and 2050, and then gravitate towards a lower value. However, to model such a more realistic forecast one would need a higher dimensional demographic model, which would complicate the model significantly, without adding additional insights. Our assumption is in line, though, with official predictions for the very long run.

Our dataset does not contain values of the pension payout, only the years when pension reforms were enacted. This implies that we can freely transform P as long as jumps are preserved. Moreover, $V(\cdot,\cdot,\cdot,\cdot)$, $U(\cdot,\cdot,\cdot)$ and K are in utility terms, so we can also freely use affine transformations, without changing the optimal timing of reforms in the model. Under Assumptions 1 and 2, we can write instantaneous utility as

$$U(P,Y,B) = u\left(Y\tilde{P}(1,B)/P\right).$$

By construction, $u(\cdot)$ has a maximum at 1. $\tilde{P}(1,B)$ is positive and decreasing in B, so a natural first-order approximation is $\tilde{P}(1,B) \approx Ae^{-cB}$, where $A > 0$ is a scaling factor and $c > 0$ measures the sensitivity of the optimal payout $\tilde{P}$ with respect to fluctuations in B. Under Assumption 3, the unconditional variance of $\log Y$ is $\sigma_y^2 \equiv \sigma_\varepsilon^2/\left(1-\phi_y^2\right)$. Now, define $y \equiv \frac{1}{\sigma_y}\log(Y)$, $p \equiv \frac{1}{\sigma_y}\log[P/A]$. $k \equiv -\frac{1}{2}u''(1)\sigma_y^2 K$, and $\gamma \equiv c/\sigma_y$. Then, we can rewrite the full optimization problem (4) up to a second-order approximation as the following minimization problem which has the same timing of reform measures:

$$v(p,y,B,b) = \min_{p'}(y-\gamma B-p')^2 + I(p'-p)k + \pi E[v(p',y',B',b')|y,B,b] \tag{8}$$

$$y' = \phi_y y + \xi_t, \qquad \xi_t \sim N\left(0,1-\phi_y^2\right) \tag{9}$$

$$B' = \lambda B + (1-\lambda)b \tag{10}$$

$$Prob\left(b' = b^j \middle| b = b^i\right) = p_b^{ij} \tag{11}$$

5.3 Calibration

We discretize the business cycle indicator y to mimic the AR(1) process as closely as possible using 201 grid points. These grid points and the corresponding transition probabilities are determined using the Rouwenhorst (1995) method (see Kopecky and Sueren, 2010, for a formal analysis). For the

AR-coefficient ϕ_y, we use the typical value used in the business cycle literature of 0.8 per quarter, which translates to 0.41 per year.

The discount factor π captures time discounting and the probability of the current government losing power. One period corresponds to one year, so we set time discounting to 0.98, corresponding to a time preference rate of about 2% a year, which is in the ballpark range of the literature. The probability of losing power is very country specific. We use the data of Armingeon *et al.* (2015) to calculate the yearly probability of an ideological change.[19] This probability varies between 0 and 50 percent. The probability of losing power clearly dominates the traditional time discounting. In our baseline simulations we use a value of 0.75 for π. Sensitivity analysis shows that our results are robust, even for extreme values of 0.5 and 0.9.

To calibrate the demographic processes, we choose the minimum and maximum values of the 25 year ahead old-age dependency ratios. The minimum value in our data is 13% (Japan, 1970) and the maximum value is 63% (also Japan, 2014). To capture all the fluctuations of this projected old age dependency ratio, we divide the assumed 10 – 70% range into 6 equally sized bins of 10 percentage points each. This yields 6 central points ranging from 15% to 65%, which we use as the grid points for our long-run old-age dependency ratio b. In our dataset, the forecast of the old age dependency ratio changes bins in 8% of the years, so we set the diagonal values of the transition matrix to 92% and all off-diagonal values to 8/5%. Our results are robust to alternative divisions of the remaining 8% over the off-diagonal values. For the current old-age dependency ratio we use an evenly spaced grid of 21 points ranging from 15% to 65%. The adjustment parameter λ is set to generate a half time of 25 years, so $\lambda^{25} = 0.5$.

This leaves us with the choices of the adjustment cost parameter $k > 0$ and the effect of the old-age dependency ratio on the optimal pension benefit, as summarized by $\gamma > 0$. We can use these two parameters without a clear empirical counterpart to match observed frequencies of the pension reforms and estimated coefficients in our empirical analysis.

Graphs 3 and 4 show the sensitivity of our model simulations to various values of k and γ. We solve the model using the structural parameters and simulate the economy starting in the unconditionally expected state and with the corresponding optimal pension payout in that state, i.e. the one that minimizes the first term on the right-hand side of (8). We then simulate 100,000 hypothetical periods and in every period check whether it is optimal to implement a contracting or expanding reform. If it is optimal to implement a reform, we set the pension benefit to the new

[19] We take the variable $CABINET\_INDEX_{it}$, which measures the ideological composition of the government through the percentages of total cabinet posts held by right-, center- and left-wing parties. The probability of an ideological change is the fraction of years that this variable changes by at least 10 percent. It ranges from zero for Switzerland to around 50 percent for Italy.

optimal one given the current state and continue the simulation. Using the simulated data and the corresponding optimal reform decisions, we perform the usual logit regressions with the change in the business cycle indicator, the change of the current old-age dependency ratio and the change in the long-term projected old-age dependency ratio (and an intercept) as explanatory variables. The adjustment cost parameter varies between 0.2 and 10. The relevance of the current old-age dependency ratio as represented by γ ranges from 0 to 10. These intervals for k and γ are wide enough to capture the observed probabilities of a reform, and the *z*-values reported earlier for the regressions based on the actual data.

Figure 3 shows the probability of a reform on the vertical axis. The reform probability exhibits hardly any sensitivity to variations in γ. Hence, the reform probability is (almost) completely determined by the fixed adjustment cost. This implies that we can always use the fixed adjustment cost to match the observed probability of a reform.

Figure 4 shows the *z*-value of the current old-age dependency ratio in the simulated logit regressions for contractionary reforms, in which the pension benefit is reduced. The graph for expansionary reforms is the mirror image of this graph. The *z*-value is mostly determined by γ, and for values of γ smaller than approximately 3, this coefficient is not significantly different from zero. The standard deviation of *B* implied by our Markov process is 0.09, so with a value of γ of 3, a one standard deviation fluctuation of *B* has the same immediate welfare loss as 0.27 standard deviations of y (which by construction has a standard deviation of 1).

5.4 Model fit

We fix all parameters to their baseline values, except the adjustment cost parameter. This adjustment cost parameter is closely related to the number of reform measures in each country, which varies widely across the countries. In our sample period, France implemented at least one reform measure in 24 out of the 44 years (55% of all years), while Iceland merely implemented a reform measure in 14% of all years. We capture this heterogeneity using a country-specific fixed adjustment cost. We set the adjustment cost parameter for each country such that the probability of a reform in our simulated data of 100,000 draws matches the fraction of years in which a country actually implemented a reform measure. This leads to two potential errors: the actual economic developments are not properly matched by our simulated data and some actual reform measures in our dataset belong to one reform "package", but are implemented in different years. The bias of the first source of errors is unclear; the second source will lead us to take a too low value of k.

To determine the model fit we compare the timing and direction of the actual reform regimes to the timing and direction as predicted by our theoretical model. We solve our model for the baseline parameters and for each of the country-specific adjustment cost parameters. As the actual business cycle indicator y we use the first principal component of the three time series of economic growth, unemployment and the budget deficit of the country. We start the simulation with the optimal pension system implied by the actual economic and demographic situation in the first sample year. Then, for every year, we check whether the economic and demographic situation in the actual data changes enough to make it optimal to change the pension payout. If this is the case, then we choose the new optimal pension payout which leads to either a predicted contracting reform or a predicted expanding reform.

Table 8 reports how often a predicted reform coincides with an actual reform in the correct direction, using various time windows to capture potential political delays. We see that the model correctly predicts an expanding reform for 19.2% of the actual country-year combinations featuring a regime with an expanding reform measure, i.e. an “Expanding only” or an “Expanding and contracting” regime. Similarly, the model correctly predicts a contracting reform for 22.0% of the actual country-year combinations with a “Contracting only” or “Expanding and contracting” regime. In reality, due to legislative lags it may take some time before a planned reform measure actually materializes. Therefore, Table 8 also reports the fraction of correct predictions if we allow for the possibility of a one- or two-year lead of the model prediction of a reform relative to an actual reform regime. In other words, in the next-to-last column the model prediction is considered correct if it takes place one year before or in the same year as the actual reform regime (and is in the correct direction as described above). If we allow for a two-year lead of the model, it correctly predicts 49.2% of the actual regimes with an expanding reform measure and 54.1% of the regimes with a contracting reform measure.

The crucial question is how good the model is if we compare its performance to a situation in which reforms occur randomly. Hence, we use a Monte Carlo simulation to compare the number of correct reforms using our model to the number of correct reforms when reforms are scattered randomly in our sample. To prevent a bias strengthening our model, in our simulation we randomly assign reforms to country-year combinations, such that for each country the number of years with a random reform coincides with the actual number of years in which a reform regime occurred in that specific country, as in the calibration of the adjustment cost. Each of the randomly assigned reforms has a 50% probability of being of the expanding type and a 50% probability of being of the contracting type. For every draw in our simulation, we determine for each element of the given set of actual country-year combinations with a reform regime whether it is predicted by a randomly assigned reform with a correct direction. We use 100.000 draws to give us the probability density function of the number of correct predictions by our randomly-assigned reforms. The values between brackets in Table 8 report the fractions of the draws in which the number of correct predictions under the random reform allocation is at least as high as under our theoretical model.

For example, when the window is confined to the same year, in about one-third of the draws does the random assignment of reforms do at least as well as our model in correctly predicting a regime with an expanding reform measure. If we allow for implementation lags, so leading predictions, the performance of our model, compared to random draws, increases substantially. The Monte Carlo analysis shows that the model especially adds predictive power for regimes with a contracting reform measure. If reforms are randomly allocated, the probability of outperforming the model, i.e. of correctly predicting at least 54.1% of the regimes with a contracting measure (with a possible lead of two years) is merely 1.6%.

6 Concluding remarks

In this paper, we empirically explored the determinants 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 in the period since the start of the seventies. Reform measures included both expansionary and contractionary measures. The determinants that we considered comprised demographic, economic, budgetary, political and crisis variables, as well as the supranational budgetary constraints on the EU countries. Crucially, we measured the determinants of reform measures in “real time”. That is, we tried to explain reform decisions on the basis of information that was available at the time the reform decision was made. Also crucially, we included the *changes* in the current and projected old-age dependency ratios, as only changes can explain why a reform could occur now if it had not occurred earlier.

We found no effect of current and projected demographic changes on the propensity to undertake reform measures. By contrast, we found quite substantial evidence that the current state of the economy, broadly defined, is the main driver of both expansionary and contractionary reform measures. We constructed a simple theoretical model with an adjustment cost of changing the pension arrangement that can account for the responsiveness to the business cycle and the non-responsiveness to demographic forecasts.

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. Our results suggest that the effectiveness of these attempts may be highly dependent on the underlying state of the countries’ economies and that these efforts are therefore best concentrated in times when their economies are in a downturn and their public budgets are under pressure. The 2005 reform of the EU Stability and Growth pact, however, allows countries to temporarily deviate from their (path to the) medium term objective due to costs of pension reform measures. However, such a deviation is not allowed from the 3%

deficit limit, which is most likely to be binding during a recession. Hence, Eurozone member states are incentivized to implement their pension reforms outside recessions, which is precisely when they are, according to our results, less inclined to implement reforms that enhance the long-run sustainability of the pension system.

References

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

Alesina, A. (1992). Political Models of Macroeconomic Policy and Fiscal Reform. *Policy Research Working Paper Series*, No.970, The World Bank.

Alesina, A. and N. Roubini (1992). Political Cycles in OECD Economies. *Review of Economic Studies* 59, 4, 663-688.

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

Armingeon, K., Isler, C., Knöpfel, L., Weisstanner, D. and S. Engler (2015). *Comparative Political Data Set 1960-2013*. 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 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.

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. 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.

Duval, R. and J. Elmeskov (2006). The Effects of EMU on Structural Reform in Labour and Product Markets, *ECB Working Paper Series*, No. 596.

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

Fernandez, R. and D. Rodrik (1991). Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty, *American Economic Review* 81, 1146-1155.

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.

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.

Häusermann, S. (2010). *The Politics of Welfare State Reform in Continental Europe: Modernization in Hard Times*, Cambridge University Press, Cambridge, UK.

ILO (2014). *International Labour Organization: NATLEX: Database of National Labour, Social Security and Related Human Rights Legislations*. Retrieved from Old-age, Invalidity and Survivors Benefits:

http://www.ilo.org/dyn/natlex/natlex4.listResults?p_lang=en&p_count=98746&p_classification=15.02&p_classcount=6430.

IMF (2004). Fostering Structural Reforms in Industrial Countries. *World Economic Outlook*, Chapter 3, April.

IMF (2014). *World Economic and Financial Surveys: World Economic Outlook Database*. Retrieved from https://www.imf.org/external/pubs/ft/weo/2015/02/weodata/index.aspx.

ISSA (2014). *International Social Security Association: Social Security Country Profiles*. Retrieved from http://www.issa.int/country-profiles

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.

Kopecky, K.A. and R.M.H. Suen (2010). Finite State Markov-Chain Approximations to highly Persistent Processes, *Review of Economic Dynamics* 13, 701--714.

Laeven, L. and F. Valencia, (2012). Systemic Banking Crises Database: An Update, *IMF Working Paper*, No. WP/12/163.

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 (2017). Do Crises Induce Reform? A Critical Review of Conception, Methodology and Empirical Evidence of the 'Crisis Hypothesis', *Journal of Economic Surveys*, forthcoming.

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

OECD (1997). *OECD Economic Surveys: Belgium/Luxembourg 1997*, OECD, Paris.

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

OECD (2013). *Pensions at a Glance 2013: OECD and G20 Indicators*. OECD Publishing, Paris.

OECD (2014). OECD Economic Outlook. Retrieved from http://stats.oecd.org/#

OECD (2014). OECD National Accounts. Retrieved from http://stats.oecd.org/#

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.

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

Ciurila, N. and W. Romp (2015). The Political Arrangement of Pay-as-You-Go Pension Systems in the Presence of Financial and Demographic Shocks, *Netspar Discussion Paper*. No. 12/2015-037.

Rouwenhorst, K.G. (1995). Asset pricing implications of equilibrium business cycle models, in Cooley, T.F., ed., *Frontiers of Business Cycle Research*, 294-330, Princeton University Press, Princeton, NJ.

Spruk, R. and M. Verbic (undated), Political Economy of Pension Reform: An Empirical Investigation. *Mimeo*, Utrecht University and University of Ljubljana.

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.

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

World Bank. (2014). *World Databank: World Development Indicators*. Retrieved from http://databank.worldbank.org/data/reports.aspx?source=2&Topic=3#

Tables

Table 1: Number of reform measures by country and category

	1. Coverage	2. Generosity and adequacy	3. Financial and fiscal sustainability	4. Work incentives
Australia	11	8	10	7
Austria	2	5	3	2
Belgium	6	14	3	3
Canada	6	10	4	2
Denmark	7	8	11	4
Finland	6	10	6	4
France	10	22	15	7
Germany	4	17	11	2
Greece	6	12	20	6
Iceland	5	1	0	1
Ireland	10	13	7	3
Italy	3	5	14	3
Japan	10	16	13	0
Luxembourg	6	12	8	2
Netherlands, the	2	7	7	2
New Zealand	5	1	9	2
Norway	1	5	3	2
Portugal	5	12	13	5
Spain	5	18	7	8
Sweden	2	8	6	4
Switzerland	7	6	13	2
United Kingdom	6	9	10	2
United States	5	12	18	1

Table 2: Number of (country, year) - combinations by category and reform regime

	1970-2013	1970-1991	1992-2013
Coverage (1)	110	52	58
Generosity and adequacy (2)	166	79	87
Expanding (1+2)	244	115	129
“Expanding only” regime	184	101	83
Financial and fiscal sustainability (3)	133	32	101
Work incentives (4)	63	7	56
Contracting (3+4)	159	36	123
“Contracting only” regime	99	22	77
“Expanding and contracting” regime	60	14	46
Total (1+2+3+4)	343	137	206

Note: “Expanding (1+2)” reports the number of different (country, year) - combinations with one or more expansionary measures. Because there are (country, year) - combinations with both a “Coverage (1)” and “Generosity and adequacy (2)” measure, the sum of “Coverage (1)” and “Generosity and adequacy (2)” exceeds “Expanding (1+2)”. Analogous for “Contracting (3+4)” and “Total (1+2+3+4)”.

Table 3: Numbers of combinations of expanding and contracting measures

	Financial and fiscal sustainability	Work incentives
Coverage	23	12
Generosity and Adequacy	36	23

Note: this table reports the numbers of times specific combinations of expanding and contracting measures are undertaken in the same year.

Table 4: Logit estimations for the baseline regressions

Independent variables	(1) Expanding Only	(2) Contracting Only	(3) Expanding and Contracting
ΔOAD	0.272	-0.184	0.387
	(0.78)	(-0.24)	(0.61)
$\Delta OAD25$	0.163	0.006	-0.112
	(1.07)	(0.02)	(-0.38)
$GROWTH$	-0.033	-0.184**	0.178
	(-0.62)	(-2.21)	(1.57)
DEF	-0.040	-0.051	0.230**
	(-0.86)	(-0.64)	(2.56)
$UNEMPL$	0.048	0.133	-0.032
	(0.97)	(1.63)	(-0.37)
D_y	-0.522	0.848	1.743**
	(-1.30)	(1.49)	(2.56)
$D_y * \Delta OAD$	-0.420	0.704	-0.322
	(-0.92)	(0.83)	(-0.41)
$D_y * \Delta OAD25$	-0.302*	-0.033	0.196
	(-1.67)	(-0.11)	(0.60)
$D_y * GROWTH$	0.193***	0.040	-0.137
	(2.63)	(0.40)	(-1.00)
$D_y * DEF$	0.072	0.026	-0.118
	(1.40)	(0.31)	(-1.25)
$D_y * UNEMPL$	-0.026	-0.002	0.049
	(-0.48)	(-0.02)	(0.51)
Wald test for significance of coefficients in 2nd sub-period			
ΔOAD	0.24	2.11	0.02
$\Delta OAD25$	2.09	0.06	0.39
$GROWTH$	9.35***	6.35**	0.27
DEF	0.84	0.34	4.75**
$UNEMPL$	0.28	5.95**	0.08
Log-likelihood	-446.1	-278.5	-187.9
N	989	989	860
LR	55.07	82.15	59.43
p-value	0.01	0.00	0.00

Notes: (i) Figures between parentheses are t-values. (ii) Further, *** denotes significance at the 1% level, ** denotes significance at the 5% level, and * denotes significance at the 10% level. (iii) $D_y = D_{1988}$ for the "Expanding only" regime, $D_y = D_{1992}$ for the "Contracting only" regime, and $D_y = D_{1997}$ for the "Expanding and contracting" regime, where D_y is a time dummy that takes a value of 0 (1) before (as of) the year y. (iv) We use the Wald test to test the individual significance of the second-period coefficients. Under the null hypothesis, the Wald test statistic, $\frac{\left(\hat{\theta}_r^j - \theta_0\right)^2}{Var\left(\hat{\theta}_r^j\right)}$, where $\hat{\theta}$ is the maximum likelihood parameter estimate under the unrestricted model and θ_0 the parameter value under the null hypothesis, converges to a chi-square distribution with one degree of freedom. In our case, $\hat{\theta}_r^j = \hat{\alpha}_r^j + \hat{\gamma}_r^j$ and $\theta_0 = 0$, where j refers to the j'th explanatory variable. (v) "N" is the number of observations, "LR" is the likelihood ratio test for the null hypothesis that the explanatory variables are jointly insignificant. It is chi-square distributed with degrees of freedom equal to the number of explanatory variables. "p-value" is the p-value for the likelihood ratio test that the explanatory variables are jointly significant.

Table 5: Mean marginal effects of the explanatory variables

	(1)	(2)	(3)
Expanding only	1970-2013	1970-1987	1988-2013
ΔOAD	0.003	0.042	-0.019
	(0.08)	(0.78)	(-0.49)
$\Delta OAD25$	-0.003	0.025	-0.018
	(-0.23)	(1.08)	(-1.45)
$GROWTH$	0.013**	-0.005	0.021***
	(2.19)	(-0.62)	(3.14)
DEF	0.001	-0.006	0.004
	(0.12)	(-0.86)	(0.92)
$UNEMPL$	0.005	0.007	0.003
	(0.89)	(0.98)	(0.53)
D_{1988}	-0.060*	-0.013	-0.091**
	(-1.89)	(-0.41)	(-2.39)
Contracting only	1970-2013	1970-1991	1992-2013
ΔOAD	0.011	-0.006	0.058
	(0.42)	(-0.24)	(1.46)
$\Delta OAD25$	-0.001	0.000	-0.003
	(-0.08)	(0.02)	(-0.25)
$GROWTH$	-0.010***	-0.006**	-0.016***
	(-3.28)	(-2.31)	(-2.56)
DEF	-0.002	-0.002	-0.003
	(-0.78)	(-0.64)	(-0.59)
$UNEMPL$	0.008**	0.004*	0.015**
	(2.40)	(1.65)	(2.44)
D_{1992}	0.064***	0.050***	0.079***
	(3.51)	(2.81)	(3.82)
Expanding and contracting	1970-2013	1970-1996	1997-2013
ΔOAD	0.010	0.009	0.006
	(0.61)	(0.61)	(0.14)
$\Delta OAD25$	-0.001	-0.003	0.008
	(-0.19)	(-0.38)	(0.62)
$GROWTH$	0.005*	0.004	0.004
	(1.71)	(1.62)	(0.52)
DEF	0.007***	0.006***	0.010*
	(3.27)	(2.60)	(2.20)
$UNEMPL$	0.000	-0.001	0.002
	(-0.19)	(-0.36)	(0.28)
D_{1997}	0.089***	0.090***	0.085***
	(4.22)	(3.56)	(4.71)

Notes: the mean marginal effects are calculated at the averages of the explanatory variables over the full sample or the indicated subsample. Further, see Notes to Table 4.

Table 6: Multinomial logit estimations with baseline specification

Independent variables	(1) Expanding Only	(2) Contracting Only	(3) Expanding and Contracting
ΔOAD	0.333	-0.097	0.433
	(0.93)	(-0.12)	(0.66)
$\Delta OAD25$	0.169	0.015	-0.126
	(1.08)	(0.05)	(-0.42)
$GROWTH$	-0.034	-0.178**	0.176
	(-0.62)	(-2.13)	(1.53)
DEF	-0.038	-0.042	0.230**
	(-0.81)	(-0.53)	(2.52)
$UNEMPL$	0.063	0.143*	0.005
	(1.25)	(1.72)	(0.05)
D_y	-0.093	0.865	1.486**
	(-0.25)	(1.48)	(2.09)
$D_y * \Delta OAD$	-0.405	0.631	-0.229
	(-0.86)	(0.73)	(-0.29)
$D_y * \Delta OAD25$	-0.300	-0.048	0.206
	(-1.61)	(-0.16)	(0.62)
$D_y * GROWTH$	0.161**	0.053	-0.158
	(2.09)	(0.52)	(-1.13)
$D_y * DEF$	0.073	0.029	-0.141
	(1.40)	(0.35)	(-1.48)
$D_y * UNEMPL$	-0.014	0.010	0.094
	(-0.25)	(0.12)	(0.93)
Wald test for significance of coefficients in 2nd sub-period			
ΔOAD	0.05	2.09	0.19
$\Delta OAD25$	1.73	0.09	0.32
$GROWTH$	5.47**	4.56**	0.05
DEF	0.92	0.09	2.72*
$UNEMPL$	1.22	7.18***	1.81
Log-likelihood	-879.41	-874.88	-873.61
N	989	989	989
LR	196.04	205.08	207.63
p-value	0.00	0.00	0.00

Notes: see Notes to Table 4. The common breakpoints in Columns (1), (2) and (3) are 1988, 1992 and 1997, respectively.

Table 7: Baseline logit regressions - 3 year averages of explanatory variables

Independent variables	(1) Expanding Only	(2) Contracting Only	(3) Expanding and Contracting
ΔOAD_3Y	-0.346	-0.005	0.877
	(-0.61)	(0.90)	(0.80)
$\Delta OAD25\_3Y$	0.470*	-1.079**	-0.210
	(1.65)	(-2.18)	(-0.39)
$GROWTH\_3Y$	0.027	-0.350**	0.037
	(0.28)	(-2.29)	(0.21)
DEF_3Y	-0.092*	-0.015	0.259**
	(-1.71)	(-0.17)	(2.55)
$UNEMPL\_3Y$	0.084	0.080	-0.035
	(1.48)	(0.88)	(-0.37)
D_y	-0.185	0.491	1.923**
	(-0.46)	(0.83)	(2.56)
$D_y * \Delta OAD\_3Y$	0.239	-0.782	-1.038
	(0.33)	(-0.62)	(-0.80)
$D_y * \Delta OAD25\_3Y$	-0.533	0.995*	0.257
	(-1.57)	(1.86)	(0.43)
$D_y * GROWTH\_3Y$	0.074	-0.005	-0.058
	(0.66)	(-0.03)	(-0.29)
$D_y * DEF\_3Y$	0.082	-0.035	-0.137
	(1.43)	(-0.38)	(-1.30)
$D_y * UNEMPL\_3Y$	-0.042	0.040	0.041
	(-0.69)	(0.42)	(0.39)
Wald test for significance of coefficients in 2nd sub-period			
ΔOAD_3Y	0.05	0.19	0.05
$\Delta OAD25\_3Y$	0.12	0.17	0.03
$GROWTH\_3Y$	2.50	18.35***	0.04
DEF_3Y	0.07	1.03	3.66*
$UNEMPL\_3Y$	0.88	3.32*	0.01
Log-likelihood	-429.2	-270.8	-183.8
N	943	943	820
LR	40.50	87.53	56.68
p-value	0.17	0.00	0.00

Notes: (i) see Notes to Table 4. (ii) The addition "_3Y" indicates that we take the three-year moving average (with the current year being the last year in the moving average).

Table 8: Fractions of country-year combinations with actual reform regimes correctly predicted by the theoretical model versus by a random allocation of reforms

	Same year	minus 1 year – same year	minus 2 years – same year
Expanding	19.2%	35.7%	49.2%
	(33.5%)	(21.2%)	(11.3%)
Contracting	22.0%	39.6%	54.1%
	(9.4%)	(4.1%)	(1.6%)

Notes: the number that is not in brackets is the fraction of actual reform regimes correctly predicted by our theoretical model, while the number within the brackets denotes the fraction of draws in which the share of correct predictions of the actual reform regimes is at least as high when reforms are randomly uniformly assigned over the sample years as under our theoretical model. A correct prediction occurs when (i) an actual "Expanding" or "Expanding and contracting" regime is matched with an expanding reform predicted by our theoretical model or the random allocation of a reform, and (ii) a "Contracting" or "Expanding and contracting" regime is matched with a contracting reform predicted by our theoretical model or the random allocation of a reform. The randomly distributed reforms are 50/50 randomly distributed as expanding or contracting.

Figures

Figure 1: Number of reform measures in different categories

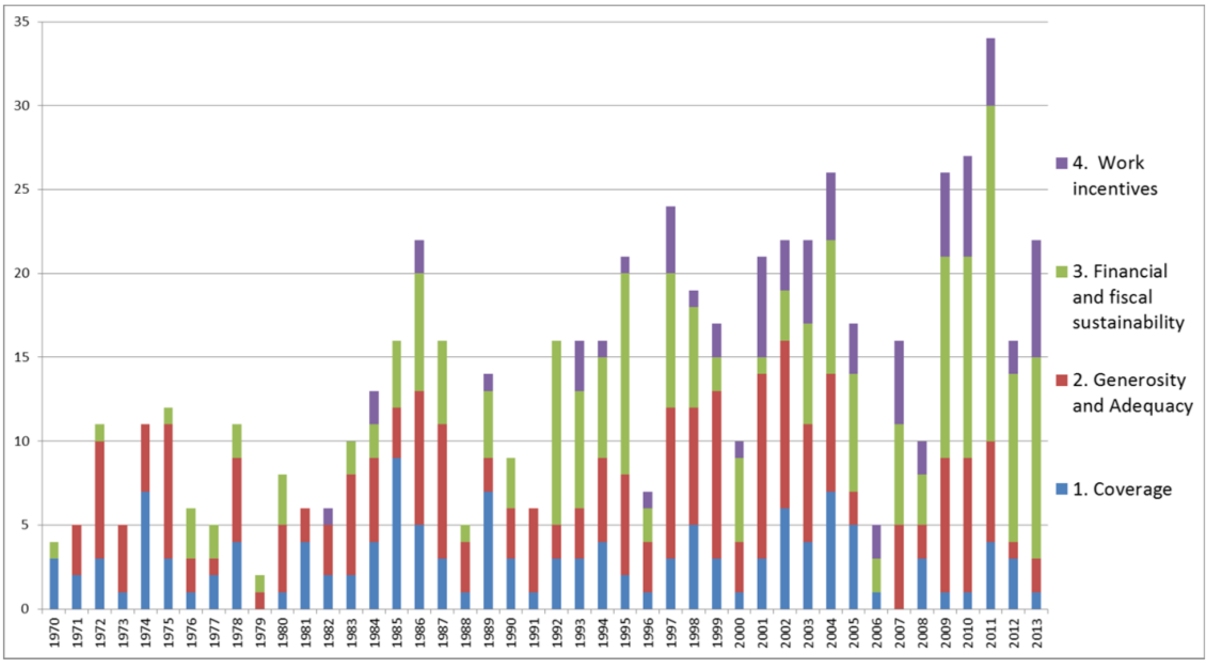
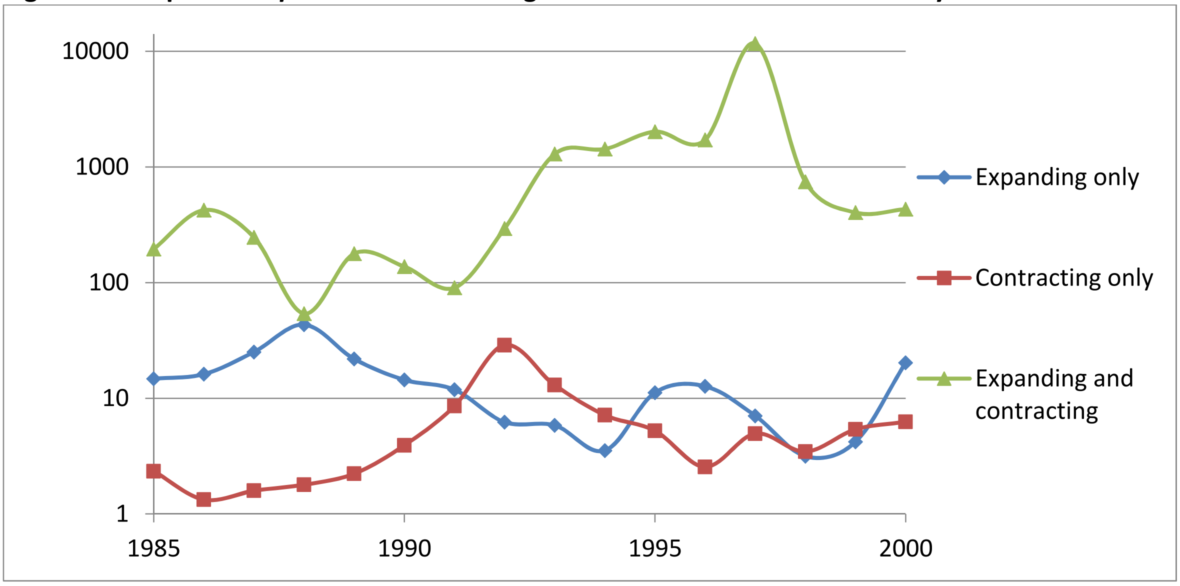


Figure 2: Reciprocal of *p*-value of test for significance structural break dummy



Note: scale is logarithmic.

Figure 3: Simulated probability of a reform (expanding or contracting) as a function of k and γ

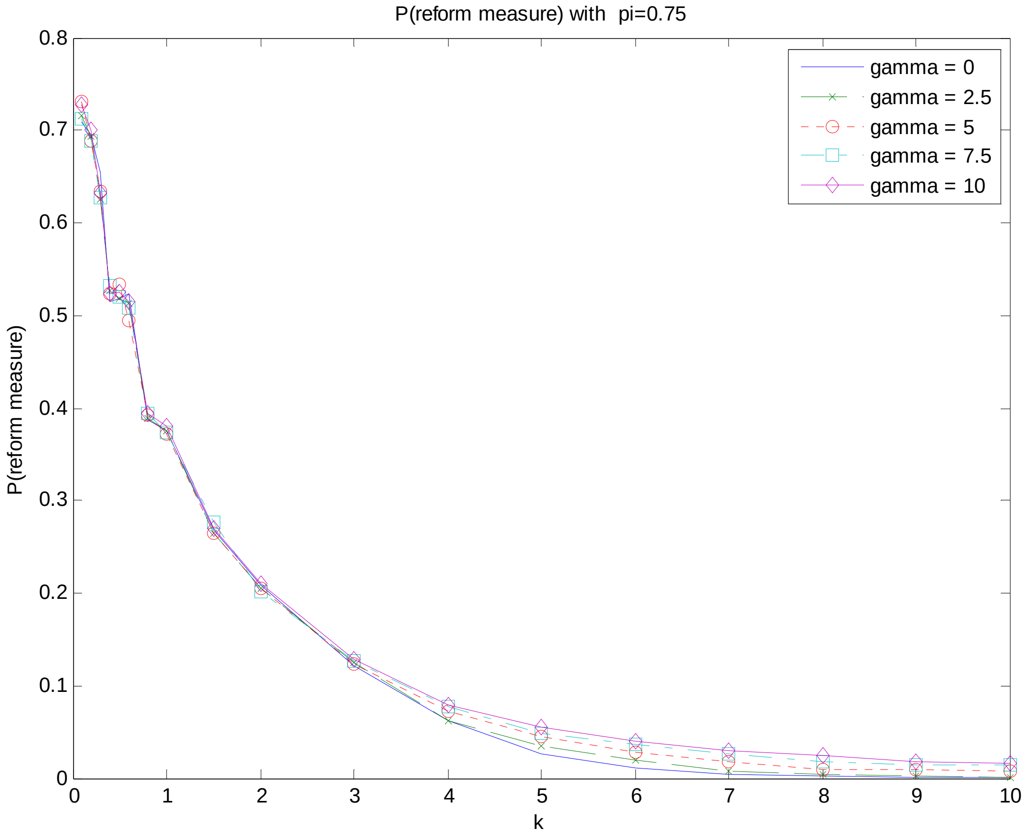
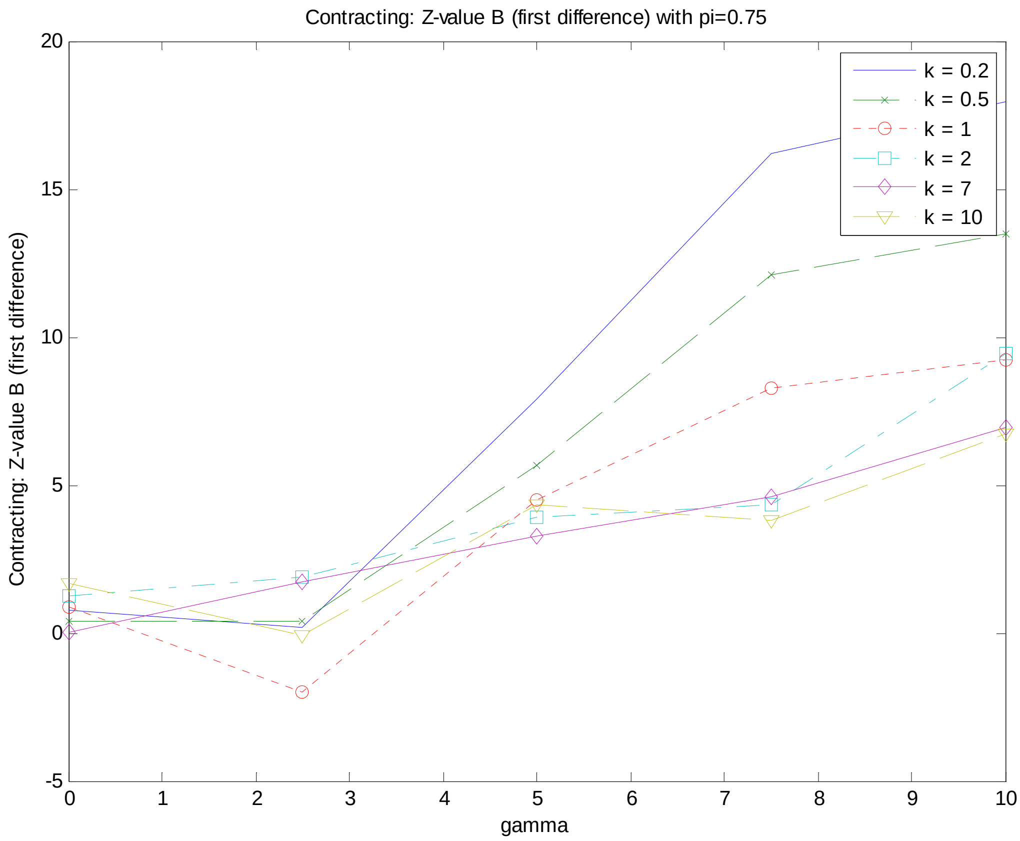


Figure 4: Simulated *z*-values of the current old-age dependency ratio as a function of k and γ



Appendices

A. Details on the collection of the reform data

A substantial amount of the input is obtained from the International Social Security Association (ISSA, 2014), which provides well-documented information on press releases and social security reforms. In addition, we make use of information from the International Labor Organization (ILO, 2014), which provides information on the legislation of old-age social security reforms.

As the date of the reform measure we always take the year in which the measure was officially legislated. This is not necessarily the year of implementation. The main reason for this is that in our further research we want to focus on what triggers pension reform measures. Hence, this requires us to focus on the information that is available at the moment that the measure is enacted. In addition, the implementation date often cannot be uniquely determined, because measures may be implemented in a number of steps. This is often the case for an increase in the statutory retirement age, which may take place in a large number of steps possibly covering a period of decades.

Reform measures can be institutional, such as the replacement of the first by a second pension pillar, or they can be incremental. The far majority of the reform measures fall into the latter category. To avoid subjective judgments, we give each reform measure the same weight in our dataset, although we know that some reform measures are more substantial than others.

Below we provide a number of examples of the classification of reform measures starting from the formulation in the original sources.

<u>Belgium 1971:</u>

ILO NATLEX (2014) writes “Royal Decree adapting certain legal provisions with the provisions of the Act of 21 December 1970 establishing a National Social Insurance Institute for freelancers.” We classify this as “Coverage”.

<u>France 1983:</u>

Gorden (1988) writes “In France, the normal retirement age under employer pensions was 65 until 1983, when all the complementary schemes based on collective agreements lowered the normal retirement age to 60 to bring them into line with the national provision approved in 1982, which introduced full career pensions at age 60 for all workers with 37.5 years of insurance.” We classify this as “Generosity and adequacy”.

<u>Belgium 2005:</u>

ISSA (2014) writes “All those who continue to work after reaching 62 will receive a supplementary pension bonus: the financial rewards will increase with the number of years worked.” We classify this as “Work incentives”.

Portugal 2005:

ISSA (2014) writes “According to the new measures the retirement age for the civil service (60 years old) will be gradually increased, by 6 months a year, until it reaches 65 years in 2015.” We classify this as “Financial and fiscal sustainability”.

Switzerland 2009:

OECD (2012) writes “Minimum rate of return on mandatory private pensions cut from 2.75% to 2% in 2009 and to 1.5% from 2012.” We classify this as “Financial and fiscal sustainability”.

Greece 2012:

OECD (2012) writes “A reduction in monthly pensions greater than €1,000 (US$1,299) by 5 per cent to 15 per cent (depending on income).” We classify this as “Financial and fiscal sustainability”.

B. Proof of Proposition 1

Define V^N and V^P as in equations (5) and (6). An upper bound for V^N is given by a situation in which the pension benefit can be changed in the next period without adjustment cost, so

$$V^N(Y,B,b) \leq \widetilde{U}(Y,B) - K + \pi E[\max_{P'} V(P',Y',B',b')|Y,B,b]. \tag{12}$$

Analogously, a lower bound for V^N is attained when the government positively changes the pension benefit both now *and* in the next period, paying the adjustment costs:

$$V^N(Y,B,b) \geq \widetilde{U}(Y,B) - (1+\pi)K + \pi E[\max_{P'} V(P',Y',B',b')|Y,B,b]. \tag{13}$$

A lower bound for V^P is given by a situation in which the government positively changes the benefit in the next period:

$$V^P(P,Y,B,b) \geq U(P,Y,B) - \pi K + \pi E[\max_{P'} V(P',Y',B',b')|Y,B,b], \tag{14}$$

while an upper bound is given by a situation in which the government can freely change the pension benefit in the next period without adjustment cost:

$$V^P(P,Y,B,b) \leq U(P,Y,B) + \pi E[\max_{P'} V(P',Y',B',b')|Y,B,b] \tag{15}$$

Subtracting (14) from (12) and combining with (12) yields

$$V^N(Y,B,b) - V^P(P,Y,B,b) \leq \widetilde{U}(Y,B) - U(P,Y,B) - (1-\pi)K = F(P,Y,B) - (1-\pi)K \leq 0,$$

hence the first result follows immediately. Subtracting (15) from (13) and combining with (13) yields

$$V^N(Y,B,b) - V^P(P,Y,B,b) \geq \widetilde{U}(Y,B) - U(P,Y,B) - (1+\pi)K = F(P,Y,B) - (1+\pi)K > 0,$$

hence the second result follows immediately as well.