

## RESEARCH ARTICLE

# The impact of product and labour market reform on growth: Evidence for OECD countries based on local projections

Jakob de Haan<sup>1,2</sup> | Rasmus Wiese<sup>1</sup>

<sup>1</sup>Faculty of Economics and Business,  
University of Groningen, Groningen, The  
Netherlands

<sup>2</sup>CESifo, Munich, Germany

## Correspondence

Rasmus Wiese, Faculty of Economics and  
Business, University of Groningen, PO  
Box 900, 9700 AV, Groningen, The  
Netherlands.

Email: [r.h.t.wiese@rug.nl](mailto:r.h.t.wiese@rug.nl)

## Summary

We examine the impact of labour and product market reforms on economic growth in 25 OECD countries between 1985 and 2013, and tests whether this impact is conditioned by the fiscal policy stance. Our local projection results suggest that controlling for endogeneity of reforms (by the Augmented Inverse Probability Weighted estimator) and fiscal policy is crucial. Our results show that product market reforms mostly cause slight negative growth, except when implemented during periods of neutral fiscal policy. Labour market reforms hurt growth under tight and neutral fiscal policy, but are conducive to economic growth if introduced during periods of expansionary fiscal policy.

## KEYWORDS

endogeneity of reforms, fiscal policy stance, local projections, structural reforms

## 1 | INTRODUCTION

Regulation is widely believed to play a role in explaining cross-country growth differences, as regulation limits the competitive pressures that challenge firms to thrive (Aghion & Griffith, 2005; Cetto et al., 2016; Nicoletti & Scarpetta, 2003; Syverson, 2011). Structural reforms are therefore often called for (cf. Draghi, 2015).<sup>1</sup>

We examine the impact of labour market and product market reforms on economic growth in 25 OECD countries. These reforms broadly involve deregulating retail trade, professional services and certain segments of network industries, primarily by reducing barriers to entry; easing hiring and dismissal regulations for regular workers; and increasing the ability of and incentives for the non-employed to find jobs (Duval & Furceri, 2018).

The effects of structural reforms on growth are important to study on its own. However, an important issue is whether reforms are better at delivering if implemented in combination with expansionary fiscal policy as suggested

<sup>1</sup>Note, however, that several studies do not find strong evidence that structural reforms enhance growth (see, for instance, Cacciatore & Fiori, 2016 and Campos & Nugent, 2018). Research based on calibrated DSGE models generally yields more favourable outcomes than cross-country studies using country level data and reform indicators (see Parlevliet et al., 2018 for a review). However, recent DSGE models that consider endogenous firm entry and search and matching frictions in the labour market suggest short-term costs from reform (Bouis et al., 2016). So, the literature is not providing a clear-cut view about the impact of reform on economic growth.

by Bordon et al. (2018) and Duval and Furceri (2018). It is likely that supportive fiscal policies enhance the positive effects of reform on economic growth (Banerji et al., 2017; IMF, 2016). Some of these reforms can entail short-term economic costs and may face political opposition. Expansionary fiscal policies may undo these negative growth effects and may support those most affected, which should help garner political support for the reform. We therefore condition the effects of reform on economic growth on the stance of fiscal policy.

We employ the local projections (LP) approach (Jordà, 2005). LP is a flexible alternative to vector autoregression models since it does not impose dynamic restrictions. Furthermore, it is better suited for estimating non-linear or state-dependent impacts, like, in our case, the stance of fiscal policy. In estimating our models, we follow Teulings and Zubanov (2014) and include the leads of the reform dummies. This approach alleviates the bias caused by overlapping forecast horizons. When calculating the forecast horizon, outcomes for observations prior to a treatment by construction overlap with the treatment ahead in time, but this is not registered in the data for the affected observation when using a standard LP set-up.

Reforms are likely to be endogenous. It is well-known that reforms are more likely to be implemented during periods of economic crises. Several previous studies do not take this endogeneity properly into account. We control for the endogeneity of reforms using the Augmented Inverse Probability Weighted (AIPW) estimator proposed by Jordà and Taylor (2015), following Glynn and Quinn (2010). We can explain the approach as follows. First, propensity scores are derived from a latent model which explains the probability of implementing a structural reform based on a number of reform predictors. Here we consider that the likelihood that a structural reform occurs may depend, among other factors, on the presence of economic crises (e.g., Duval et al., 2021) and the fiscal policy stance (Mierau et al., 2007). These propensity scores are used to correct for selection bias by reweighting the sample in the LP outcome regressions such that we achieve a quasi-random distribution of treatment and control observations. The LP model is used to estimate conditional means in the treatment group and in the control group (observations in which no reform took place) separately based on a number of determinants. Finally, the differences in weighted conditional means (where weights are represented by the inverse propensity scores of each observation) at each horizon between the treatment and control groups are computed to estimate the average treatment effects (ATEs) of reforms on economic growth.

The papers most closely related to our work are Bordon et al. (2018) and Duval and Furceri (2018). Bordon et al. (2018) investigate the impact of structural reforms on employment using OECD labour market reform indicators and the local projection approach, while controlling for endogeneity. However, unlike Bordon et al. (2018), we examine the impact of reforms on economic growth. Instead of using the OECD reform indicators, we use the reform indicators of Duval et al. (2018). According to Duval and Furceri (2018), this database identifies the exact timing of major legislative and regulatory actions by advanced economies since the early 1970s in key labour and product market policy areas. Furthermore, it captures reforms in areas for which OECD indicators exist but do not cover all relevant policy dimensions.

Duval and Furceri (2018) also use local projections and the same database as the current paper. There are three main differences with our work. First, instead of using fiscal policy shocks identified as the forecast error of government expenditure, we identify the stance of fiscal policy based on the presence of fiscal adjustments and fiscal expansions following the approach suggested by Wiese et al. (2018). In our view, fiscal adjustments or expansions better capture the stance of fiscal policy; it builds upon the literature initiated by Alesina and Perotti (1995); see also Alesina et al. (2019). Second, we control for the endogeneity of reforms. Third, we include treatment leads to alleviate the bias from overlapping forecast horizons as proposed by Teulings and Zubanov (2014).

Our findings indicate that controlling for endogeneity of reforms and the stance of fiscal policy is crucial. In contrast to the results of most previous studies, our results suggest that product market reforms mostly cause slightly negative growth. Labour market reforms hurt growth under restrictive and neutral fiscal policy but are conducive to economic growth if introduced during periods of expansionary<sup>2</sup> fiscal policy. So, our results confirm the finding of Duval and Furceri (2018) that labour reforms enhance growth only if accompanied by expansionary fiscal policy. However, we do not confirm their result that product market reforms enhance growth.

The remainder of the paper is organized as follows. Section 2 discusses the data used. Section 3 outlines our methodology, while Section 4 presents our main findings. Section 5 offers a robustness analysis, while Section 6 concludes.

<sup>2</sup>We use 'expansionary' ('restrictive') fiscal policy and 'loose' ('tight') fiscal policy interchangeably.

## 2 | DATA

### 2.1 | Structural reform

Most previous research on the impact of structural reform uses OECD regulation indicators (see, for instance, Bouis et al., 2012; Faccini, 2014; and Bordon et al., 2018). These range from 0 to 6 to capture the restrictiveness of regulation in labour and product markets. A reform is then identified as a fall of the index. In our research, we instead use the narrative reform database of Duval et al. (2018). Drawing on Duval and Furceri (2018), we may summarize the methodology used to construct this database as follows. In a first step, all legislative and regulatory actions related to product and labour market regulation mentioned in the OECD Economic Survey are identified for the 26 countries over the entire sample. Next, out of the more than 1000 actions, major reforms are identified based on three criteria: (1) the OECD Economic Survey uses strong normative language to define the action, suggestive of an important measure; (2) the policy action is mentioned repeatedly across different editions of the OECD Economic Survey for the country considered; (3) the OECD indicator of product and labour market regulation displays a very large change.

The reform variables we apply are two dummies (labour and product market reforms) equal to one if there is a major reform in a particular year identified by Duval et al. (2018). There are 248 product market reforms and 79 labour market reforms in our sample which runs from 1985 to 2013 (see Table SA1.1, which is available on the journal website). Note that the reform indicator is equal to one only if in a particular year a labour or product market reform was introduced. So, if a reform indicator is one during several subsequent years, a reform took place in each of these years.

As shown in Table SA1.1, Luxembourg and the United States implemented few product market reforms, while the remaining countries implemented a fair amount of these reforms (see also Figure SA1.1). Luxembourg and the United Kingdom did not implement labour market reforms in our sample period, while countries like Italy, Portugal, Spain and the Scandinavian countries (apart from Finland) have been most active in this area (see also Figure SA1.2). Both types of reforms are fairly evenly distributed over time.

### 2.2 | Fiscal policy stance

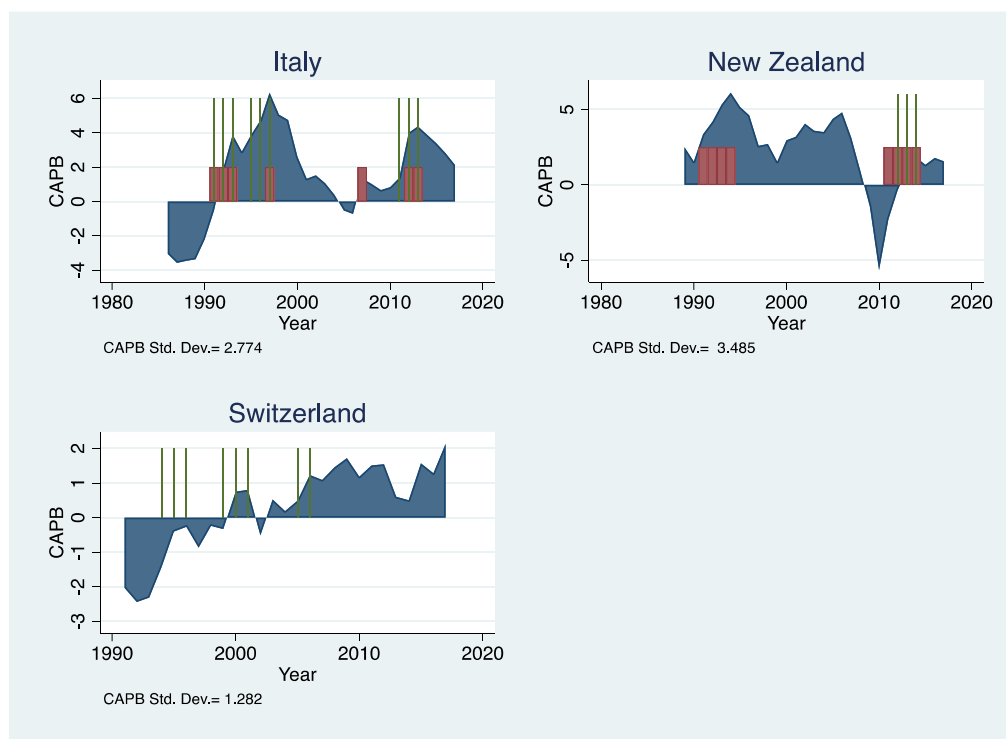
Following the approach suggested by Wiese et al. (2018), we apply the Bai and Perron (BP) (Bai & Perron, 1998; Bai & Perron, 2003) structural break filter to the cyclically adjusted primary budget balance (CAPB) as share of GDP to identify fiscal adjustments and fiscal expansions. We prefer this approach over alternatives used in the literature on fiscal adjustments for reasons explained below. Our approach to identify the beginning of a period with tight or expansionary fiscal is based on the identification of statistically significant changes in the Data Generating Process of the CAPB.<sup>3</sup> Bai and Perron (1998, 2003) developed a general method for this purpose. Consider a model with  $m$  possible structural breaks:

$$y_t = \delta_j + \mu_t \quad (t = 1, \dots, T; j = 1, \dots, m+1) \quad (1)$$

where  $y_t$  is the dependent variable, in our case, the cyclically adjusted primary budget balance in each individual country,  $\delta_j$  is a vector of estimated constants, that is, the mean of the  $m+1$  different segments of the time series  $y_t$ , and  $\mu_t$  is the error term. The BP filter generates the segmented route through the times series that yields the lowest Sum of Squared Residuals (SSR) up to a maximum number of breaks. The maximum number of breaks is restricted by the trimming parameter  $h$ , which specifies a minimum number of observations that have to occur between consecutive breaks. We have set  $h = 0.15$  which means that the time between breaking points should be not less than 15% of the time period.<sup>4</sup> The process underlying the algorithm is straightforward. First, it searches for all possible sets of breaks up to a

<sup>3</sup>This part heavily draws on Wiese et al. (2018).

<sup>4</sup>The choice of trimming parameter is based on the recommendations in Bai and Perron (1998, 2003). It is not econometrically sound to set it lower than 15% in finite samples. Furthermore, increasing the trimming percentage to  $h = 0.20$  causes only minor changes in the identified periods. Increasing the trimming percentage much further makes it increasingly difficult to identify the starting point of adjustment periods because of the restriction it places on the time that has to pass between breaking points, thus the precision of the method declines.



**FIGURE 1** Fiscal policy outcomes. *Note:* This figure shows the cyclically adjusted primary balance (CAPB) (in blue), and years identified as fiscal adjustments using two approaches, namely the requirement that the change in the CAPB is larger than 1.5 percentage points (in red bars) and the approach outlined in this section (in green spikes)

maximum restricted by the trimming parameter  $h$ , and determines the set that minimizes the SSR for each number of breaks. Then,  $F$ -tests determine whether the improved fit produced by allowing additional breaks is sufficiently large compared to what can be expected randomly on the basis of the asymptotic distribution derived in Bai and Perron (1998). We used the test procedure recommended by Bai and Perron (2003) to select the optimal number and timing of breaks. That is, dependent on properties of the individual time series, we chose the appropriate filter specification and test. Generally, though, the error distribution is allowed to differ across segments.<sup>5</sup> Autocorrelation and potential heteroskedasticity is modelled non-parametrically by running the filter using a heteroskedasticity and autocorrelation-consistent estimate of the variance-covariance matrix.

The BP method identifies the break date (fiscal adjustment or fiscal expansion initiation) as the first year after an upward or downward structural break in CAPB. Therefore, we take a 1-year lag to identify the start of the fiscal adjustment or expansion. In line with Wiese et al. (2018), we define the periods such that tight fiscal policy after an upward structural break continues as long as the change in the CAPB is positive, and expansionary fiscal policy after a downward structural break continues as long as the change in the CAPB is negative. Observations that are not classified as either expansionary or tight fiscal policy are labelled as neutral fiscal policy.

Annual data for the cyclically adjusted budget balance come from the OECD and begin in 1985 for some countries, but later for several others. Due to the limited availability of CAPB data we lose observations when we estimate models that are dependent on the state of fiscal policy.

To illustrate why we think that the one-size-fits-all approach does not properly identify fiscal adjustments, take a look at Figure 1 which shows the cyclically adjusted budget balance (as % of GDP) in Italy, New Zealand and Switzerland. The figure suggests that Switzerland did not have any fiscal adjustment according to the approach to identify a fiscal adjustment if the change in the (cyclically adjusted) budget balance exceeds some threshold. However, this outcome is deceptive as it does not take into account the low volatility of the fiscal policy process in Switzerland. Our approach takes this country-specific volatility in fiscal policy outcomes into account. There are

<sup>5</sup>This means that we do not assume that fiscal institutions are constant over time within countries.

several years during which our approach to identify fiscal adjustments based on breaks in the data series for the budget balance suggests that a fiscal adjustment occurred. The reverse holds for Italy and especially New Zealand which have higher volatility in fiscal policy outcomes. As a consequence, the one-size-fits-all approach is likely to identify more fiscal adjustments for Italy and New Zealand than for Switzerland (as illustrated by the red bars in Figure 1), simply because the change of the budget balance is the key criterion to identify a fiscal adjustment (i.e., type I errors). By the same token, the one-size-fits-all approach is less likely to detect significant changes in fiscal policy in countries where the budgetary process leads to less volatile policy outcomes, such as in Switzerland. In that case this approach suffers from type II errors.

An alternative to the one-size-fits all approach of Alesina and co-authors is the so-called actions or narrative approach (see Devries et al., 2011; Guajardo et al., 2014, for details). Drawing on Guajardo et al. (2014), we can explain it as follows. Policy documents are used to establish whether discretionary changes in tax rates and government spending were motivated by a response to the economic outlook or not. Policy changes are considered exogenous if there is no indication in those documents that the policy change is in response to prospective economic conditions. The documents considered include budgets, central bank reports, Convergence and Stability Programs submitted by the authorities to the European Commission, IMF Recent Economic Developments reports, IMF Staff Reports and OECD Economic Surveys, as well as country-specific sources. These documents provide evidence of what policymakers believed at the time when policy measures were taken, and provide estimates of the measures' budgetary impacts. Guajardo et al. (2014) focus on deficit-driven fiscal policy changes.

Although much better than the one-size-fits-all approach, the narrative approach also has some problems (Holden & Larsson Midthjell, 2013). It might be hard to assess the true intention of a policy change solely by reading policy documents. By including only actions, which are motivated by debt reduction, and thus omitting fiscal actions for stabilization purposes, one may obtain an inaccurate picture of the actual fiscal stance. Furthermore, as Guajardo et al. (2014) acknowledge, the narrative approach records changes in fiscal policy when they occur, which ignores the role of anticipation effects. In addition, Perotti (2012) provides examples showing that 'action-based' fiscal data might also be subject to important measurement errors, as governments quite often reverse announced spending cuts in subsequent supplementary budgets. Finally, although the narrative approach aims to get fiscal policy shocks that are exogenous, Brueckner (2021) reports that year-to-year changes in temperature, GDP growth of trading partners, and the international commodity price index have a significant and negative contemporaneous effect on action-based fiscal consolidations and a significant positive contemporaneous effect on annual GDP growth. Based on these results, the author concludes that it is highly unlikely that action-based fiscal consolidations are exogenous.

Table SA3.1 compares the periods of tight fiscal policy according to Guajardo et al. (2014), fiscal adjustments under the Alesina and Ardagna (2010) approach and those identified following the methodology of Wiese et al. (2018). The table shows that our approach yields fewer periods with tight fiscal policy than the narrative approach. However, many periods identified under our approach are also identified under the narrative approach. This also holds, albeit to a lesser extent, for the tight fiscal policy periods identified under the one-size-fits-all approach.

An alternative to our approach, would be to use fiscal policy shocks. This is what Duval and Furceri (2018) do. They identify these shocks as the forecast error of government consumption expenditure to GDP, following the approach used by Auerbach and Gorodnichenko (2012). The forecasts of government consumption used in the analysis are those reported in the fall issue of the OECD's Economic Outlook for the same year. In our view, this measure suffers from some shortcomings, notably that fiscal policy is more than only government consumption, and that it is assumed that only unexpected fiscal policy can have real effects.

As Figure 2 shows, product market reforms seem unrelated to fiscal policy. This is confirmed by the low correlation reported in Table 1. However, labour market reforms happen more frequently during periods of tight fiscal policy than during periods with loose fiscal policy, but this is also caused by the fact that we observe fewer fiscal expansions than fiscal adjustments.

## 2.3 | Dependent and other variables

Most other data come from the Penn World Tables (PWT) version 9.0 (Feenstra et al., 2015). The dependent variable is cumulative real GDP growth per capita projected stepwise forward in time, so 0 to 1, 0 to 2, and so on, until 0 to 5 years. The cumulative growth rates are calculated based on real growth rates (log differences of real GDP in PPP 2011 US\$, divided by populations size, both from PWT 9.0).





**FIGURE 2** Fiscal policy stance and market reforms. *Note:* The figure displays the distribution of major product and/or labour market reforms during loose (expansionary), neutral and tight (restrictive) fiscal policy. In total there are 140 observations with tight fiscal policy, 112 with loose fiscal policy and 545 with neutral fiscal policy. There are 24 observations where a product and labour market reform is introduced simultaneously

**TABLE 1** Correlation between reforms and fiscal policy stance

	Tight	Neutral	Loose
Product market reform	0.0473	−0.0732	0.0461
Labour market reform	0.1143*	−0.0576	−0.0481

\*Significance at the 5% level after Bonferroni correction.

The political variables considered in the first stage of the AIPW estimator are own updates of variables used in Wiese et al. (2018). They are explained in more detail in Section 3. In the robustness tests we consider additional variables from the OECD and the European Commission. Table SA2.1 provides a description of all variables used and their sources.

### 3 | ESTIMATION METHODS

The basic LP regression model that we estimate is:

$$\log y_{i,t+h} - \log y_{i,t} = \alpha_i + \beta_{1jh} \sum_{j=0}^5 d_{i,t-j} + \beta_{2lh} \sum_{l=0}^3 (\log y_{i,t-l} - \log y_{i,t-1-l}) + \beta_{3h} \sum_{h=1}^h d_{i,t+h} + \beta'_{4h} X_{i,t} + u_{i,t+h} \quad (2)$$

where  $h = 1, \dots, 5$  is the forecast horizon, and  $\log y_{i,t+h} - \log y_{i,t}$  denotes the cumulative growth rates of real GDP over the forecast horizon.  $\alpha_i$  denotes country fixed-effects and  $d_{i,t}$  are the reform indicators. We include treatment lags in our models. But contrary to the leads, it is an empirical issue how long the effect of reforms persists in the data. We use Akaike's information criterion to determine the lag length which consistently tells us to use 5 lags of the treatment variable.

In all the OLS LP regressions (and in the AIPW LP outcome regression, see below) we include an AR(4) term for the growth rate between  $t-l$  and  $t-1-l$ .<sup>6</sup> The term  $\beta_{3h} \sum_{h=1}^h d_{i,t+h}$  captures the Teulings and Zubanov (2014) correction. The leads are included to avoid the bias that results from overlapping forecast horizons.<sup>7</sup>  $X_{i,t}$  is a vector of additional control variables.  $X_{i,t}$  contains the output gap to control for reversion to mean growth. The output gap is calculated using the Hodrick-Prescott filter using high smoothing ( $\lambda = 100$ ) as recommended in Jordà and Taylor (2015).<sup>8</sup>  $X_{i,t}$  also includes the change in physical capital (gross investments relative to GDP) and the percentage change from year to year in the human capital index from PWT 9.0. These Solow variables control for the impact of changes in the capital stock on economic growth that may also result from reforms. Notice, that both reform indicators at time  $t$  are included simultaneously in the regressions. For example, when we estimate the effect of product market reforms  $X_{i,t}$  contains the contemporaneous labour market reform indicator, but not its lags or leads.

To estimate the responses of product and labour market reforms conditional on fiscal policy we augment Equation 2. Specifically, we interact all variables in the model with a set of dummy variables that indicate the stance of fiscal policy at the time of reform. So, we estimate state-dependent local projections with an observed threshold variable. In Equation 3 below  $I_{z,i,t}$  is the threshold variable. Specifically it is a vector which contains a set of mutually exclusive dummy variables indicating the stance of fiscal policy at the time of the reform (see Ramey & Zubairy, 2018, and Auerbach & Gorodnichenko, 2012, for a similar approach).<sup>9</sup> The  $z$  in  $I_{z,i,t}$  indicates which dummy variable in the vector is used and thus whether fiscal policy is tight, neutral or loose when the dummy variable is equal to one. Thus, we are allowing the forecast of  $\log y_{i,t+h} - \log y_{i,t}$  to differ according to the stance of fiscal policy at the time when the reform is implemented. This way, the dynamics of a different fiscal policy regimes are allowed to carry over to a new fiscal policy regime. In other words, the lags of GDP growth and the lags of reforms may carry over from periods with a different stance of fiscal policy. The coefficient and standard errors on  $d_{i,t-0}$  are then used to calculate the impulse response for each forecast horizon for each state of fiscal policy.

$$\log y_{i,t+h} - \log y_{i,t} = \sum_{z=1}^3 I_{z,i,t} \left[ \alpha_i + \beta_{1jh} \sum_{j=0}^5 d_{i,t-j} + \beta_{2lh} \sum_{l=0}^3 (\log y_{i,t-l} - \log y_{i,t-1-l}) + \beta_3 \sum_{h=1}^h d_{i,t+h} + \beta'_{4h} X_{i,t} \right] + e_{i,t+h} \quad (3)$$

The major drawback of Equations 2 and 3 is that they ignore that structural reforms may be introduced in countries/years where the expected benefits of reform are higher than in countries/years where no reforms are introduced. In other words, reforms are more likely in years where they are politically viable. Failing to account for this can lead to selection bias. Therefore, we proceed with a quasi-experimental method, namely the AIPW estimator proposed by Jordà and Taylor (2015) and Glynn and Quinn (2010). In the first step, we estimate logit models to estimate the probability that a structural reform occurs. This latent variable framework captures the idea that reforms are introduced in periods where they are politically viable. In the second step, we use local projections specified as Equations 2 and 3, but weighing observations inversely according to the predicted probabilities from the logit model.

Specifically, observations in which a reform took place are assigned a weight,  $w$ , by the inverse of  $p$ , the probability score ( $w=1/p$ ), whereas the observations without reform receive a weight of the inverse of one minus the probability score ( $w=1/(1-p)$ ). This means that treated observations with a low probability score receive a higher weight in the regression along with control observations with a high probability score. This places more weight on observations that are comparable and hence reduces selection bias. The augmented weighting adds an adjustment factor to the treatment effect when the estimated probability scores are close to zero or one. The method is doubly robust and only requires one of the following two conditions to hold: The conditional mean model is correctly specified or the probability score model is correctly specified. Weighting can be interpreted as removing the correlation between the covariates and the

<sup>6</sup>Again, we Akaike's information criterion to determine the lag length. Our choice of information criteria is based on the fact that the Akaike criterion is least likely to choose an autoregressive process of too low order; see Lütkepohl and Kratzig (2004).

<sup>7</sup>The bias increases with the forecast horizon, see Teulings and Zubanov (2014). The leads of the treatment dummies ensure that it is registered in the data if the outcome for a specific observation is affected by a treatment ahead in time. This most often is the case for control observations, that is, country-year pairs where no reform took place. However, reforms may occur repeatedly within our forecast horizon of 5 years. In that case, the Teulings and Zubanov (2014) approach also registers that the outcome of a treated observation may be affected by later treatments, which otherwise would have meant an upward bias in the effect of reforms.

<sup>8</sup>Our results are robust to using OECD data on the output gap which is calculated using a production function approach; the correlation coefficient between the OECD output gap and our own is 0.845.

<sup>9</sup>Note that such a fully specified interaction model is equivalent to a set-up where the sample is split according to the stance of fiscal policy.

reform indicator, and regression removes the direct effect of the covariates (see Imbens & Wooldridge, 2009, for more details). We report the Average Treatment Effect (ATE), which is calculated as the average difference between treated and non-treated (control) observations based on the weighted OLS regression line for both groups.

To correct for the imported uncertainty from the first stage propensity score estimation in the second stage, we calculate block-bootstrapped standard errors in our AIPW models. That is, we construct the bootstrap by repeatedly drawing blocks of observations, that is, drawing countries rather than individual observations with replacement. This way, serial correlation in the error terms is also taken into account. First, we test whether spatial dependence is present in the disturbances between the cross-sectional units when using standard errors clustered at the country level. For this purpose, we use the Pesaran (2015) test, which is standard normally distributed. So, a value of the test statistic outside the  $[-1.96, 1.96]$  interval rejects the null hypothesis of weak cross-sectional dependence. Although the tests sometimes reject the hypothesis, we use the cluster-bootstrapped errors since cross-sectional dependence does not bias our point estimates; it only leads to an efficiency loss, (see Elhorst, 2013). In the one-stage simple LP results we calculate Spatial Correlation Consistent (SCC) standard errors as proposed by Driscoll and Kraay (1998).

## 4 | MAIN RESULTS

### 4.1 | Local projection results

Table 2 shows the LP estimates of our basic model (Equation 2). So, here we do not control for endogeneity and do not condition on the stance of fiscal policy. Figure 3 shows the corresponding impulse response functions for product and labour market reforms, based on the estimates shown in Table 2.

It is quite remarkable that in this very simple set-up, product and labour market reforms hardly affect output growth. The estimated coefficient on the output gap coefficient is negative (as expected). This means that on the upturn of the business cycle growth in the future is predicted to be significantly negative and vice versa.

The sum of the coefficients on the lagged dependent variable in column (1) is larger than one, which may imply a non-stationary growth process. However, panel stationarity tests reject non-stationarity (results available on request). In columns (2)–(5) of Table 2, the sum is much larger than one, but that is not a surprise since we estimate cumulative growth rates. The insignificant physical and human capital elasticities are perhaps not surprising in a sample of OECD countries; see Mankiw et al. (1992) for a similar finding.

Next, we condition on the stance of fiscal policy. Table 3 presents the estimation results and Figure 4 shows the corresponding impulse response functions. Surprisingly, the main takeaway from Table 3 is that product and labour market reforms only have an effect after 5 years during periods of tight fiscal policy. Specifically, for product market reforms the accumulative effect on growth after 5 years is 1.9% of additional GDP (see column 5 in the row tight fiscal policy). For labour market reforms the effect is significant and positive, after 5 years 3.7% of additional GDP is estimated.

### 4.2 | Quasi-experimental results

In an ideal Randomized Controlled Trial (RTC) setting where treatments are assigned randomly, we would expect the probability density function for each control variable included in Equations 2 and 3 to be the same for each subpopulation of treated and control units. The overlap of the densities should be close to perfect. For example, the distribution of the deviation between actual GDP and potential GDP (the output gap) should be similar for the subpopulation where a major product market reform takes place and the subpopulation of all other (control) observations. A simple way to check whether this condition holds is to do a test of equality of means between the subsamples. This is done in Table 4a. As evident, especially the balance of the output gap between treated and control observations is a cause of concern. This is an indication that we cannot assume that treatments are assigned randomly as is done in the simple LP analysis above. Specifically, the balance tests in Table 4a, 4b indicate that the output gap on average is negative (implying that the economy is running below its potential) for treated observations compared to control observations. This suggests that labour and product market reforms cannot be viewed as exogenous events.

Notice that this balance condition is also behind the implicit assumption that we can estimate the simple LPs presented above by restricting the coefficients of the controls in Equations 2 and 3 to be the same for the treatment and the control groups. The AIPW estimates below relaxes this assumption, as a regression is specified separately for both the



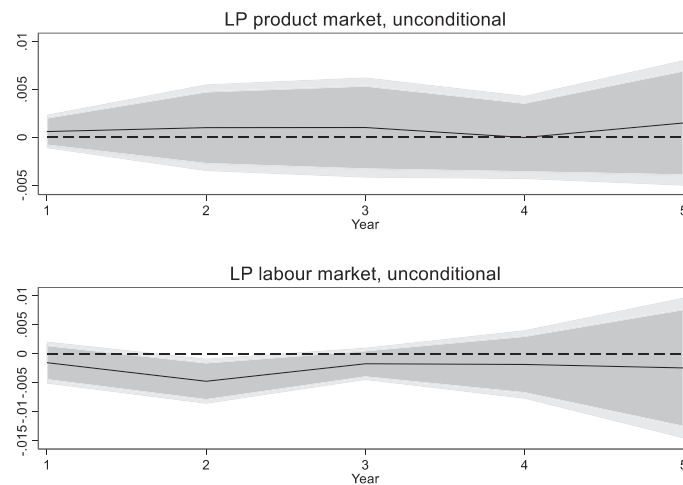
**TABLE 2** Local projection estimates of the effect of product and labour market reforms in  $t = 0$  on cumulative economic growth in  $t = 1, \dots, 5$ 

<b>Panel A: Product market reforms</b>					
<b>VARIABLES</b>	<b>(1) Year 1</b>	<b>(2) Year 2</b>	<b>(3) Year 3</b>	<b>(4) Year 4</b>	<b>(5) Year 5</b>
Product market reform	0.001 (0.001)	0.001 (0.002)	0.001 (0.003)	−0.000 (0.002)	0.001 (0.003)
Output gap	−0.095*** (0.010)	−0.207*** (0.012)	−0.291*** (0.010)	−0.352*** (0.012)	−0.392*** (0.017)
d.y	0.662*** (0.103)	0.933*** (0.145)	1.075*** (0.139)	1.040*** (0.129)	0.776*** (0.214)
d.y <sub>t-1</sub>	0.158*** (0.038)	0.558*** (0.049)	0.817*** (0.092)	0.885*** (0.080)	1.070*** (0.138)
d.y <sub>t-2</sub>	0.245*** (0.037)	0.389*** (0.058)	0.463*** (0.062)	0.646*** (0.067)	0.778*** (0.156)
d.y <sub>t-3</sub>	0.113** (0.044)	0.287*** (0.066)	0.504*** (0.090)	0.666*** (0.142)	0.748*** (0.149)
Gross capital formation	0.099 (0.080)	0.138 (0.151)	0.025 (0.146)	−0.195 (0.183)	−0.301 (0.227)
Change in human capital index	0.120 (0.193)	0.202 (0.324)	−0.033 (0.385)	0.202 (0.440)	0.551 (0.432)
Labour market reform	−0.001 (0.002)	−0.004* (0.002)	−0.000 (0.002)	−0.001 (0.003)	0.000 (0.005)
Constant	−0.005** (0.003)	−0.007 (0.005)	0.000 (0.007)	0.009 (0.008)	0.018* (0.010)
R-squared	0.626	0.800	0.878	0.879	0.836
Pesaran CD-test statistic	37.66	35.05	31.07	27.87	24.20
<b>Panel B: Labour market reforms</b>					
<b>VARIABLES</b>	<b>(1) Year 1</b>	<b>(2) Year 2</b>	<b>(3) Year 3</b>	<b>(4) Year 4</b>	<b>(5) Year 5</b>
Labour market reform	−0.002 (0.002)	−0.005** (0.002)	−0.002 (0.001)	−0.002 (0.003)	−0.002 (0.006)
Output gap	−0.093*** (0.010)	−0.203*** (0.013)	−0.289*** (0.011)	−0.356*** (0.014)	−0.403*** (0.022)
d.y	0.654*** (0.097)	0.908*** (0.129)	1.046*** (0.126)	1.027*** (0.112)	0.802*** (0.188)
d.y <sub>t-1</sub>	0.151*** (0.039)	0.549*** (0.045)	0.804*** (0.092)	0.890*** (0.080)	1.068*** (0.175)
d.y <sub>t-2</sub>	0.248*** (0.037)	0.383*** (0.055)	0.462*** (0.057)	0.649*** (0.087)	0.772*** (0.186)
d.y <sub>t-3</sub>	0.105** (0.039)	0.278*** (0.063)	0.498*** (0.099)	0.648*** (0.150)	0.733*** (0.166)
Gross capital formation	0.098 (0.068)	0.127 (0.128)	−0.002 (0.134)	−0.211 (0.172)	−0.272 (0.197)
Change in human capital index	0.055 (0.226)	0.097 (0.392)	0.053 (0.427)	0.600 (0.501)	1.506** (0.639)
Product market reform	0.000 (0.001)	0.001 (0.003)	0.001 (0.002)	−0.000 (0.002)	0.001 (0.003)
Constant	−0.004* (0.002)	−0.005 (0.005)	0.001 (0.006)	0.007 (0.008)	0.015 (0.014)
R-squared	0.625	0.800	0.880	0.877	0.821
Pesaran CD-test statistic	38.27	35.45	31.13	30.45	32.22
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes
Observations	690	665	640	615	590
Number of countries	25	25	25	25	25

Note: The table shows the local projection estimates of product market reforms (panel A) and labour market reforms (panel B) on cumulative economic growth, unconditional of the fiscal policy stance. The models are based on Equation 2, and include treatment leads equal to the forecast horizon, five lags of the treatment variable, and country fixed-effects. The number of treatment lags was determined by Akaike's information criterion. The Pesaran (2015) test for cross-sectional dependence is conducted on the disturbances from a model with SEs clustered at the country level. The tests indicate spatial dependence in the errors. Spatial correlation consistent standard errors are shown in parentheses.

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.1$ .

treatment group and the control group (Imbens & Wooldridge, 2009; Jordà & Taylor, 2015). The difference in the predicted outcomes of  $\log y_{i,t+h} - \log y_{i,t}$  between each regression for the treatment and the control group then serves to calculate the (weighted) ATEs.



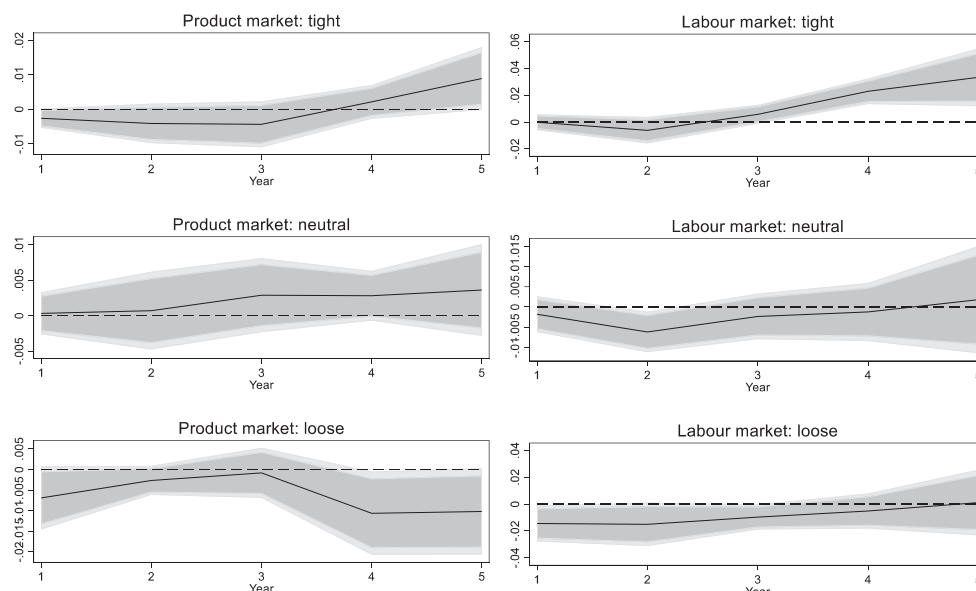
**FIGURE 3** Impulse responses of local projection estimates of the effect of product and labour market reforms on cumulative economic growth in  $t = 1, \dots, 5$ . *Note:* The solid black lines in the figure plots the impulse responses of product market (upper panel) and labour market (lower panel) reforms on cumulative economic growth based on the regressions shown in Table 2. So, the position of the line at  $t = 5$  shows the accumulated impact of reform on GDP in percentages, 5 years after the reform. The y axis displays single-digit growth rates at the second decimal place. The dark grey shaded areas display the 90% SCC error bands; the light grey shaded areas display the 95% SCC error bands. For both product and labour market reforms the impact is not significant. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

**TABLE 3** Local projection estimates of the effect of product and labour market reforms in  $t = 0$  on cumulative economic growth in  $t = 1, \dots, 5$ , dependent on the stance of fiscal policy

VARIABLES	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Product market reforms</b>					
Tight	−0.003* (0.001)	−0.004 (0.003)	−0.004 (0.003)	0.002 (0.002)	0.009* (0.005)
Neutral	0.000 (0.002)	0.001 (0.003)	0.003 (0.003)	0.003 (0.002)	0.004 (0.003)
Loose	−0.007* (0.004)	−0.003 (0.002)	−0.001 (0.003)	−0.011* (0.005)	−0.010* (0.005)
Number of obs. Product market	643	618	593	568	543
R-squared	0.673	0.826	0.897	0.900	0.870
Pesaran CD test statistics	37.66	35.05	31.07	27.87	24.20
<b>Labour market reforms</b>					
Tight	−0.000 (0.003)	−0.006 (0.005)	0.006 (0.004)	0.023*** (0.005)	0.033*** (0.011)
Neutral	−0.002 (0.002)	−0.006** (0.003)	−0.002 (0.003)	−0.001 (0.004)	0.002 (0.007)
Loose	−0.014** (0.007)	−0.015* (0.008)	−0.010* (0.005)	−0.005 (0.007)	0.001 (0.013)
Number of obs. Labour market	643	618	593	568	543
R-squared	0.669	0.822	0.895	0.899	0.861
Pesaran CD test statistics	38.27	35.45	31.13	30.45	32.22
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

*Note:* The table shows the local projection estimates of labour and product market reforms on cumulative economic growth, conditional of the fiscal policy stance. The models are based on Equation 3, and include treatment leads equal to the forecast horizon, five lags of the treatment variable, and country fixed-effects. The number of treatment lags was determined by Akaike's information criterion. The Pesaran (2015) test for cross-sectional dependence is conducted on the disturbances from a model with SEs clustered at the country level. The tests indicate spatial dependence in the errors during periods of neutral fiscal policy. Spatial correlation consistent standard errors are shown in parentheses.

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.1$ .



**FIGURE 4** Impulse responses of local projection estimates of the effect of product and labour market reforms on cumulative economic growth in  $t = 1, \dots, 5$ , dependent on the stance of fiscal policy. *Note:* The figure plots the impulse responses of product market and labour market reforms on cumulative economic growth based on the regressions in Table 3 with the solid black lines. So, the position of the line at  $t = 5$  shows the accumulated impact of reform on GDP in percentages, 5 years after the reform. The y axis displays single-digit growth rates at the second decimal place. The impulse responses are partitioned on fiscal policy stance. The dark grey shaded areas display the 90% SCC error bands; the light grey shaded areas display the 95% SCC error bands. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

**TABLE 4a** Balance of covariates in treatment and control group

	Difference in means (treated minus control)
<b>Product market reform</b>	
Output gap	−0.0276* (0.0163)
d.y	−0.0031* (0.0020)
Gross capital formation	−0.0013 (0.0018)
Change in human capital index	0.0001 (0.0002)
<b>Labour market reform</b>	
Output gap	−0.0551** (0.0274)
d.y	−0.0045 (0.0035)
Gross capital formation	0.0020 (0.0032)
Change in human capital index	0.0006 (0.0004)
Observations	691

*Note:* Standard errors of a two-sided  $t$ -test are reported in parentheses.

\*\* $p < 0.05$ . \* $p < 0.1$ .

When policy interventions like labour and product market reforms are driven by endogenous responses to control variables (as shown in Table 4a), the observed treatment and control units can be viewed as being oversampled from the part of the distribution in which the propensity score of treatment reaches high values. The simple local projections presented above are based on the sampled distribution and will therefore be biased. Too much weight is given to treated observations with a high probability of treatment and too little weight is given to control observations with a high probability of treatment. Inverse weighting using propensity scores shift the probability mass away from the oversampled region of the distribution towards the under-sampled region. This shift rebalances the sample such that we can view

TABLE 4b Balance of covariates in treatment and control group conditional on fiscal stance

	Difference in means (treated minus control)		
	Tight	Neutral	Loose
<b>Product market reform</b>			
Output gap	0.0110 (0.0317)	−0.0538*** (0.0192)	0.0487 (0.0526)
d.y	0.0111** (0.0047)	0.0017 (0.0026)	0.0092 (0.0057)
Gross capital formation	−0.0091* (0.0047)	0.0006 (0.0023)	−0.0075 (0.0054)
Change in human capital index	−0.0011** (0.0005)	0.0002 (0.0004)	0.0000 (0.0006)
<b>Labour market reform</b>			
Output gap	−0.0978** (0.0464)	−0.0170 (0.0356)	−0.0069 (0.0772)
d.y	−0.0143 (0.0091)	−0.0023 (0.0036)	0.0118 (0.0114)
Gross capital formation	0.0124 (0.0086)	−0.0016 (0.0031)	−0.0044 (0.0096)
Change in human capital index	0.0003 (0.0007)	0.0001 (0.0005)	0.0037* (0.0018)
Observations	104	430	109

Note: Standard errors of a two-sided *t*-test are reported in parentheses.

\*\*\**p* < 0.01. \*\**p* < 0.05. \**p* < 0.1.

the re-weighted sample as reconstructing the true distribution of outcomes under treated and control observations. In other words, we can view the rebalancing as if we had observed a random sample for each group, unaffected by endogenous responses to control variables. Thus, the regression for both the control group and the treatment group are less susceptible to bias and their difference can be used to calculate an unbiased estimated of the ATE of reforms on economic growth (see Imbens & Wooldridge, 2009; Jordà & Taylor, 2015, for more details).

To estimate the propensity scores, ideally any predictor of treatment should be included regardless of whether that variable is included in the model specified in Equations 2 and 3. Therefore, we follow Jordà and Taylor (2015) and estimate a saturated propensity score model. As predictors of the likelihood of structural reform ('the treatment in  $t+1$ '), we use the following variables. First, the output gap and the lagged growth rates as these variables capture the idea that reforms are more likely to occur after times of economic crisis (Drazen & Grilli, 1993). In line with this argument, we also include the unemployment rate and the inflation rate, as these variables may also signify difficult economic times. We also add the fiscal policy stance as the likelihood of reforms may differ across different types of fiscal policy regimes. We add political variables to account for the fact that reforms are more viable under certain political circumstances. Specifically, we add the following: (1) A variable counting the number of years a government has held office to capture the idea that reforms become less likely the longer a government holds office. (2) An election variable reflecting that an executive or legislative election took place to capture the idea that reforms typically are more likely after a new government takes office (Haggard & Webb, 1993). (3) A variable measuring government ideology to capture the idea that the political colour of a government matters in terms of the policies it implements (Hibbs, 1977). (4) A variable measuring political fractionalization to capture the idea that more politically fragmented governments may find it harder to implement economic reforms (Alesina & Drazen, 1991). We also control for the possibility that labour and product market reforms may be related (Fiori et al., 2012) by including labour market reforms as predictor of product market reforms in  $t + 1$ , and vice versa for labour market reforms. In the logit model for labour market reforms, we also add institutional variables capturing the strictness of hiring and firing conditions for workers on temporary or regular contracts. This takes a level effect into account, as countries with very flexible hiring and firing conditions are typically less likely to reform the labour market (Turrini et al., 2015).<sup>10</sup> We also include variables for duration dependence (Carter & Signorino, 2010). Specifically, we add a variable that counts the number of years since the last reforms, plus its squared and cubed term. *F*-tests show that the duration dependence variables are jointly significant and therefore should be in the model. Table SA2.1 provides a description of the variables used.

<sup>10</sup>This set of variables is in line with the robust triggers of labour and product market reforms identified in Duval et al. (2021).

TABLE 5 Logit regressions predicting the probability of treatment at  $t + 1$ , marginal effects

VARIABLES	(1)	(2)	(3)
	Product market reform	Labour market reform	Joint reform
Output gap	−0.688*** (0.113)	0.332*** (0.101)	−0.016 (0.120)
d.y	6.561*** (1.507)	−2.547* (1.486)	−1.185 (2.871)
d.y <sub>t-1</sub>	1.222 (0.890)	−0.244 (0.666)	−0.492 (0.711)
d.y <sub>t-2</sub>	2.450*** (0.826)	−0.100 (0.646)	1.841* (0.987)
Gross capital formation	3.079** (1.552)	−1.171 (1.503)	−1.716 (2.949)
Change in human capital index	−3.909 (7.610)	15.666** (6.353)	10.126 (9.512)
Labour market reform	0.102** (0.049)		−0.002 (0.034)
Product market reform		0.033 (0.027)	0.000 (0.026)
Fiscal expansion dummy	0.033 (0.048)	−0.021 (0.039)	−0.002 (0.044)
Fiscal adjustment dummy	0.029 (0.044)	0.045 (0.032)	−0.004 (0.030)
Unemployment rate	−0.011 (0.008)	0.034*** (0.008)	0.006 (0.006)
Inflation rate	−0.007 (0.007)	−0.011 (0.007)	0.004 (0.011)
Government yrs. in office	0.008 (0.006)	−0.018*** (0.006)	−0.011* (0.006)
Any election (executive and/or legislative)	0.011 (0.034)	0.024 (0.026)	−0.011 (0.027)
Government ideology	0.039* (0.021)	0.022 (0.018)	0.030 (0.019)
Political fragmentation	−0.190*** (0.073)	−0.016 (0.058)	−0.034 (0.055)
Strictness of employment protection, temporary employment		0.025 (0.022)	−0.019 (0.032)
Strictness of employment protection, regular employment		0.131* (0.071)	0.034 (0.066)
Third degree polynomial of time since reform	Yes	Yes	Yes
Country Fixed-effects	Yes	Yes	Yes
Observations	710	591	353
Area under ROC curve	0.789	0.849	0.913

Note: The table reports the marginal effects at the means of a logit specification to predict the probability of treatment in  $t + 1$ . In model 3 treatment is defined as observations in which both a product market reform and a labour market reform occurred simultaneously, there are 24 treatments in that case. As a consequence, we can only use the 13 countries in which reforms occurred at least once simultaneously to estimate the model, due to the inclusion of fixed effects. Standard errors clustered at the country level are shown in parentheses.

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.1$ .

An important thing to note here is that although relatively few variables are highly significant, the model has a high predictive ability: the ‘area under the ROC curve’ is above or close to 0.8 in all the reported logit models.<sup>11</sup> In all specifications shown in Table 5, the area under the ROC curve is statistically significantly different from 0.5.<sup>12</sup>

Figures 5–7 provide smooth kernel density estimates of the distribution of the propensity scores for treatment and control units to check for overlap. The plotted densities are based on models 1–3 in Table 5, respectively. In the ideal RCT setting, the overlap between the distribution of propensity scores for treated and control units would be near identical. Although the logit models used to estimate the propensity scores all have high predictive ability, Figures 5–7 make clear that we have considerable overlap between the distributions for treated and control units. This indicates that we

<sup>11</sup>ROC stands for Receiver Operating Characteristics. It is also referred to as the Correct Classification Frontier. If the model had no predictive ability, the area under the ROC curve would be 0.5. A perfect classification ability would correspond to an area under the ROC curve equal to 1. The area under the ROC curve has an approximate normal distribution in large samples.

<sup>12</sup>In line with Jordà and Taylor (2015), we include country-dummies in the estimations. If we estimate the models in Table 5 without fixed effects (FE) the predictive ability declines, but the area under the ROC curve is still statistically different from 0.5. The models with country fixed effects turned out to be superior in predicting treatment. So, we proceed with the FE specification for the propensity scores in the AIPW estimates regardless of the incidental parameter problem in the logit model.



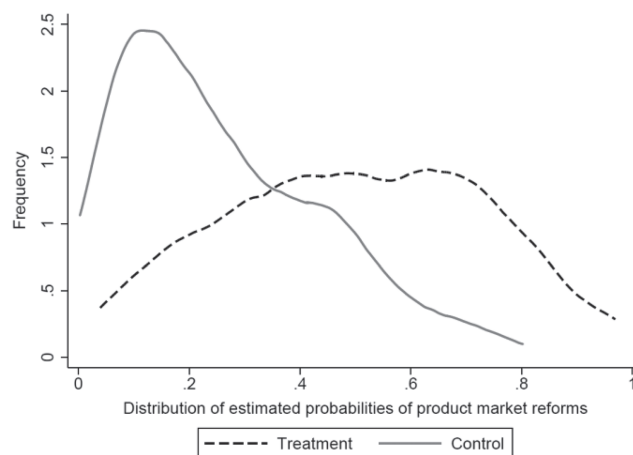


FIGURE 5 Overlap of propensity scores for product market reforms

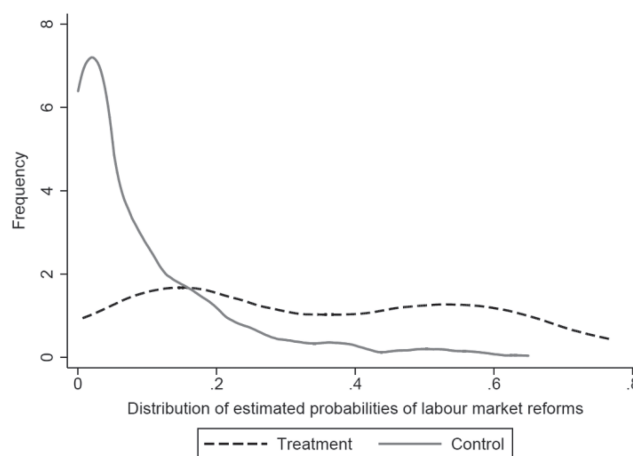


FIGURE 6 Overlap of propensity scores for labour market reforms

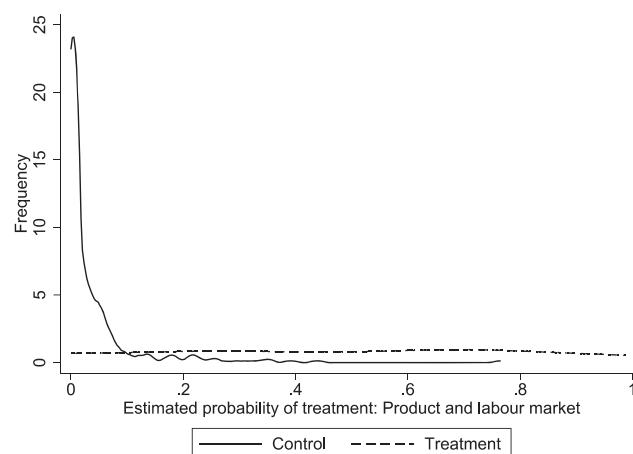


FIGURE 7 Overlap of propensity scores for the interaction of labour and product market reforms. *Note:* In the AIPW estimates of the interaction of labour and product market reforms below we truncate propensity scores at 0.05 for p-scores lower than 0.05 due to observations with a very low p-score. Otherwise the estimator becomes unstable (1 divided with a very small number will give a very large weight to treated observation with a low predicted p-score). To keep symmetry, we also truncate at high propensity scores, so above 0.95. As Figure SA5.1 in the Figure SA5.1 shows, there is variability in the propensity scores for control observations. This is masked by the smooth kernel density plot in Figure 6 because we have many observations with a propensity score close to zero

TABLE 6 Average treatment effect of reforms, AIPW estimates

VARIABLES	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Product market reform</b>	0.001 (0.002)	0.001 (0.002)	0.003* (0.002)	0.002 (0.002)	0.000 (0.003)
Pesaran CD test statistics	1.10	5.31	4.46	0.87	1.67
Observations	619	594	569	544	519
<b>Labour market reform</b>	0.002 (0.002)	−0.003 (0.002)	−0.007*** (0.002)	−0.006*** (0.002)	−0.008*** (0.003)
Pesaran CD test statistics	5.88	3.51	2.83	3.26	4.76
Observations	541	518	495	472	449
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

Note: The table shows the ATE responses of the AIPW local projection estimates of labour and product market reforms on cumulative economic growth, unconditional of fiscal policy stance. Compared to Table 1, the number of observations drops due to unavailability of some of the reform predictor variables. The models are based on Equation 2. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses:

\* $p < 0.10$ . \*\*\* $p < 0.01$ .

have a satisfactory logit model that can be used to identify the ATEs properly using our quasi-experimental estimation strategy.

Figures 6 and 7 also make clear that some treated units have a propensity score that is very close to 0. In practice, this means that these observations get very high weights when weighing inversely with the propensity score. Although the AIPW estimator adds an adjustment factor to the treatment effect when the estimated probability scores are close to 0 for treated observations (and close to 1 for control observations) this is not enough to stabilize the estimator concerning joint reforms. Therefore, we truncate the propensity scores for joint reforms (see the notes to Figure 7), following Imbens (2004) and Cole and Hernan (2008).

Tables 6 and 7 present the results using the quasi-experimental doubly-robust AIPW estimator proposed by Jordà and Taylor (2015). Table 6 shows the estimation results if we do not condition on fiscal policy, while Figure 8 offers the corresponding impulse response functions. If we do not condition on the fiscal policy stance at the time of the reform, the effects of reform on GDP growth are small. Only labour market reforms affect economic growth: after 4 years, GDP has declined by 0.6%.

However, when we estimate state dependent ATEs for each type of reform during different types of fiscal stances a more fine-grained pattern emerges (Table 7 and Figure 9). Product market reforms mostly cause slight negative growth, except for a very small positive effect during periods of neutral fiscal policy. Labour market reforms hurt growth if introduced during periods of tight fiscal and neutral fiscal policy, but they are conducive to economic growth if introduced during periods of loose fiscal policy.

As shown, our findings change drastically when we control for the fact that the assignment of treatment is non-random. Unconditional of fiscal policy, product market reforms have no statistically significant effect on economic growth, while the unconditional effect of labour market reforms is significantly negative throughout the evaluation period.

When conditioning on fiscal policy a more fine-grained pattern emerges. Contrary to the simple LP results where treatment selection is ignored, we now find mostly negative effects of both product and labour market reforms during fiscal adjustments. However, labour market reforms have a positive effect on growth if implemented during a fiscal expansion, while their effect on growth is again negative if implemented when fiscal policy is neutral. Product market reforms have little to no effect regardless of the fiscal policy stance.

## 5 | ROBUSTNESS ANALYSIS

We check whether our main AIPW findings are sensitive to the way we identify the fiscal policy stance. As an alternative to the method based on structural break tests, we apply threshold criteria as usual in the literature on fiscal

TABLE 7 Average treatment effect of reforms dependent on the stance of fiscal policy, AIPW estimates

VARIABLES	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Product market reforms</b>					
Tight	0.003 (0.003)	0.003 (0.003)	0.001 (0.003)	−0.007 (0.005)	−0.014*** (0.005)
Neutral	0.000 (0.003)	0.002 (0.003)	0.006*** (0.002)	0.005* (0.003)	0.003 (0.004)
Loose	−0.004 (0.003)	−0.005** (0.002)	−0.003 (0.004)	−0.012** (0.005)	−0.010 (0.007)
Number of obs. Product market	619	594	569	544	519
Pesaran CD test statistics	4.33	4.83	5.20	0.53	1.87
<b>Labour market reforms</b>					
Tight	−0.022*** (0.005)	−0.022*** (0.007)	−0.078** (0.034)	−0.011 (0.012)	−0.024** (0.010)
Neutral	0.001 (0.001)	−0.005** (0.002)	−0.007*** (0.002)	−0.005** (0.002)	−0.006 (0.003)
Loose	0.007* (0.003)	0.041** (0.019)	0.035*** (0.010)	0.013*** (0.005)	0.036*** (0.009)
Number of obs. Labour market	541	518	495	472	449
Pesaran CD test statistics	1.42	3.32	0.33	3.69	4.69
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

Note: The table shows the ATE responses of the AIPW local projection estimates of labour and product market reforms on cumulative economic growth, conditional of fiscal policy stance. The models are based on Equation 3. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.

\* $p < 0.10$ . \*\* $p < 0.05$ . \*\*\* $p < 0.01$ .

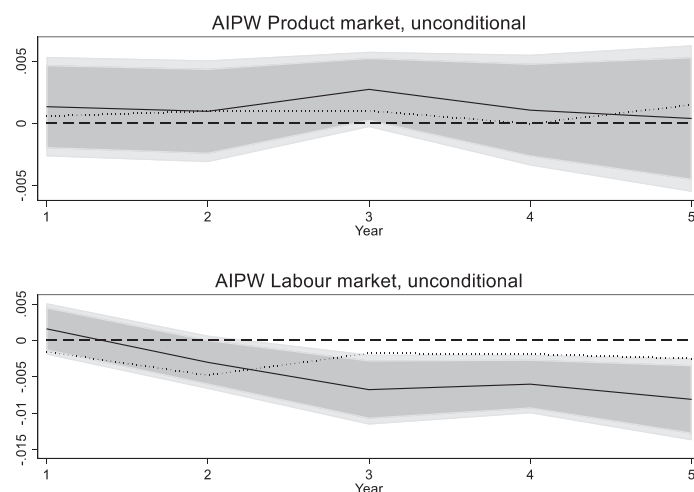
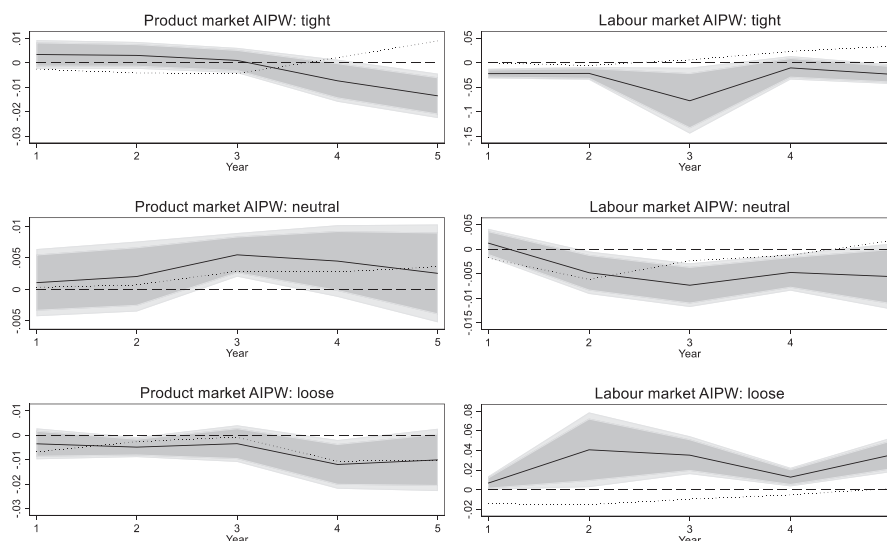


FIGURE 8 Average treatment effect (ATE) responses of Augmented Inverse Probability Weighted (AIPW) local projection estimates of the effect of product and labour market reforms on cumulative economic growth in  $t = 1, \dots, 5$ . Note: The figure plots the ATE responses of product (upper panel) and labour market (lower panel) reforms on cumulative economic growth based on the regressions in Table 6 with the solid black lines. So, the position of the line at  $t = 5$  shows the accumulated impact of reform on GDP in percentages, 5 years after the reform. The y axis displays single-digit growth rates at the second decimal place. The dark grey shaded areas display the 90% error bands; the light grey shaded areas display the 95% error bands. The dotted line displays impulse responses from the simple LP regressions. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

adjustments (see Section 2 for a discussion). Specifically, we define the start of a fiscal adjustment as a positive change in the CAPB larger than 1.5% of GDP; the adjustment continues as long as the change in CAPB is positive. A negative change in CAPB smaller than −1.5% of GDP indicates the start of a fiscal expansion, which continues as long as the



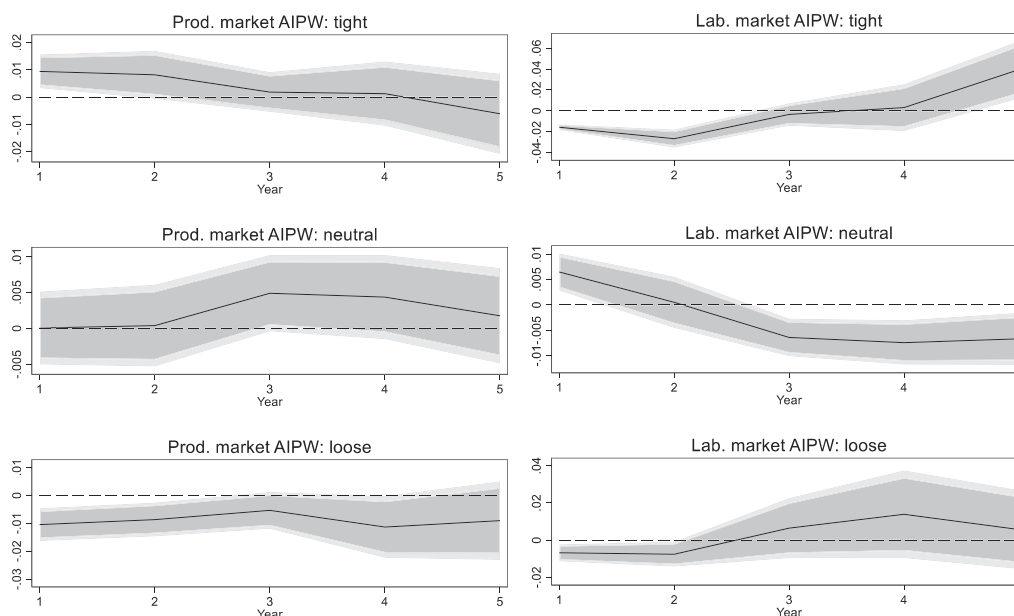
**FIGURE 9** Average treatment effect (ATE) responses of Augmented Inverse Probability Weighted (AIPW) local projection estimates of the effect of product and labour market reforms on cumulative economic growth in  $t = 1, \dots, 5$ , dependent on the stance of fiscal policy. *Note:* The figure plots the ATE responses of product market (left panel) and labour market (right panel) reforms on cumulative economic growth based on the regressions in Table 7 with the solid black lines. The impulse responses are partitioned by fiscal policy stance, the upper panels display the ATE responses under tight fiscal policy, the middle panels under neutral fiscal policy and the lower panels under loose fiscal policy. The y-axis displays single-digit growth rates at the second decimal place. The dark grey shaded areas display the 90% error bands; the light grey shaded areas display the 95% error bands. The dotted line displays impulse responses from the conditional LP regressions. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

**TABLE 8** Average treatment effect of reforms dependent on the stance of fiscal policy according to Alesina–Ardagna criterion, AIPW estimates

VARIABLES	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Product market reforms</b>					
Tight	0.009*** (0.003)	0.008* (0.005)	0.002 (0.004)	0.001 (0.006)	−0.006 (0.008)
Neutral	0.000 (0.003)	0.000 (0.003)	0.005* (0.003)	0.004 (0.003)	0.002 (0.003)
Loose	−0.011*** (0.003)	−0.009*** (0.003)	−0.005 (0.003)	−0.011** (0.006)	−0.009 (0.007)
Number of obs. Product market	619	594	569	544	519
Pesaran CD test statistics	3.30	4.98	4.49	0.15	1.18
<b>Labour market reforms</b>					
Tight	−0.016*** (0.001)	−0.027*** (0.004)	−0.004 (0.006)	0.003 (0.012)	0.039*** (0.014)
Neutral	0.006*** (0.002)	0.000 (0.003)	−0.006*** (0.002)	−0.007*** (0.002)	−0.007** (0.003)
Loose	−0.007*** (0.002)	−0.008** (0.003)	0.006 (0.008)	0.014 (0.012)	0.006 (0.011)
Number of obs. Labour market	545	522	499	476	453
Pesaran CD test statistics	9.38	8.16	2.56	1.17	4.95
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

*Note:* The table shows the ATE responses of the AIPW local projection estimates of labour and product market reforms on cumulative economic growth, conditional of fiscal policy stance. Fiscal policy stance is identified using the Alesina–Ardagna criterion (see Section 2). The models are based on Equation 3. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.

\* $p < 0.10$ . \*\* $p < 0.05$ . \*\*\* $p < 0.01$ .



**FIGURE 10** Average treatment effect (ATE) responses of Augmented Inverse Probability Weighted (AIPW) local projection estimates of the effect of product and labour market reforms on cumulative economic growth in  $t = 1, \dots, 5$ , dependent on the stance of fiscal policy according to Alesina-Ardagna criterion. *Note:* The figure plots the ATE responses of product market (left panel) and labour market (right panel) reforms on cumulative economic growth based on regressions in Table 8 with the solid black lines. The impulse responses are partitioned by the fiscal policy stance determined using a threshold approach. The upper panels display the ATE responses under tight fiscal policy, the middle panels under neutral fiscal policy and the lower panels under loose fiscal policy. The y axis displays single-digit growth rates at the second decimal place. The dark grey shaded areas display the 90% error bands; the light grey shaded areas display the 95% error bands. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

change in the CAPB is negative. That way, we identify 152 periods with fiscal adjustments, 128 periods with fiscal expansions and 517 periods with neutral fiscal policy; see Figure SA1.3 for the distribution of reforms over these types of fiscal policy stances. Table SA3.1 provides a detailed comparison of fiscal adjustments under both approaches. Table 8 presents the estimation results and Figure 10 shows the impulse response functions.

Figure 10 suggests that our conclusion that labour market reforms enhance economic growth under loose fiscal policy also holds under the alternative definitions of the fiscal policy stance. Under tight and neutral fiscal policy, labour market reforms have a negative effect on growth; under tight fiscal policy this negative effect is significant only in the first years after the reform, while under neutral fiscal policy it becomes significant after some years. In line with our previous findings, product market reforms generally have a negative or non-significant effect on economic growth.

As an additional robustness test, we add the debt burden when predicting the weights in the logit model. The debt burden is measured as the product of the debt-to-GDP ratio and the short-term interest rate. This variable significantly predicts product market reforms, but does not predict labour market reforms.<sup>13</sup> As can be seen in Figures SA4.1 and SA4.2, our main findings are robust to this change in the logit model specification.

Furthermore, the European output gap is added in both the logit model and the outcome regressions (Equations 2 and 3). Firstly, reforms may be imposed supra-nationally at the European level when the EU experiences a (severe) downturn in the business cycle, that is, an economic crisis. As European business cycles may comove (cf. Mink et al., 2012), we control for a reversal to the European mean GDP growth in the outcome regressions. As can be seen in Figures SA4.3 and SA4.4, the results are largely robust to these additions to our models. Product market reforms have hardly any effect on GDP growth in European Union member countries,<sup>14</sup> while labour market reforms are conducive to economic growth if implemented during periods of loose fiscal policy.

<sup>13</sup>We lose observations when including the debt burden in our logit models. Specifically, we end up with 606 observations for product market reforms and 522 observations for labour market reforms in the logit models. For product market reforms, the US is dropped from the sample due to the fixed-effects and missing observations on the debt burden.

<sup>14</sup>Controlling for the EU output gap has the consequence that sample is reduced to include only European Union member countries, and only in the time frame in which they were members.



Additionally, we analyse the joint effect of labour and product market reforms. In practice, that amounts to analysing whether reforms work better or worse when implemented as broad reform packages, that is, simultaneous reforms in both the product and labour market. Unfortunately, we only have 25 observations in which major reforms occur in both the product and labour market simultaneously. Therefore, it is not possible to conduct this analysis while conditioning on fiscal policy. There are simply too few treated observations in each cell for the different types of fiscal policy.

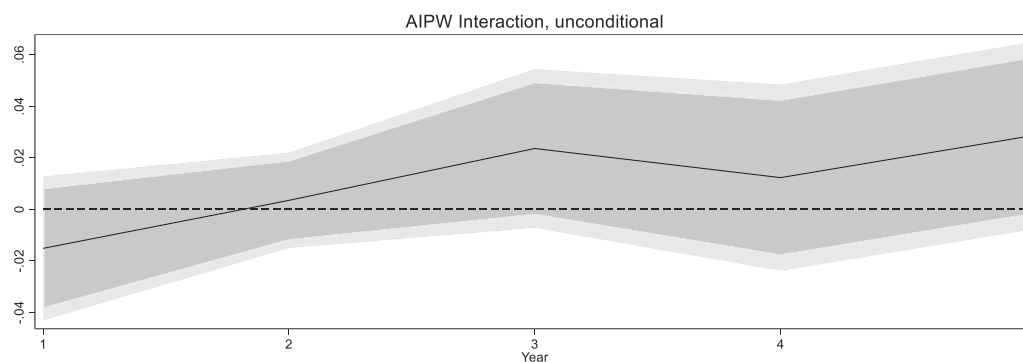
The results reported in Table 9 and Figure 11 suggest that the initial effect of joint labour and product market reforms is negative in the short term but in the medium term the effect becomes positive. This conclusion follows from the fact that in the short run the effect of economy wide reforms is negative but insignificant. After 2 years, the effect becomes positive and after 5 years GDP has grown by 3%, although marginally insignificant.

To investigate whether the main AIPW results are merely artefacts of the estimator applied to our data, we conduct falsification tests by simulating all our main AIPW regression specifications with placebo reforms. To maintain comparability with our main results, we randomly draw the placebo reforms from a binomial distribution with a probability of treatment equal to the proportion of product and respectively labour market reforms in our sample, see Borjas (2017) for a similar approach. In Figures SA4.5 and SA4.6, we report the results of the

**TABLE 9** Average treatment effect of the interaction of labour and product market reforms, AIPW estimates

	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Reform interaction effect</b>	−0.015 (0.014)	0.003 (0.009)	0.024 (0.016)	0.012 (0.018)	0.028 (0.019)
Obs.	327	314	301	288	275
Pesaran CD test statistics	0.22	1.77	1.14	−1.31	−0.50
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

*Note:* The table shows the ATE responses of the AIPW estimates of labour and product market reforms occurring simultaneously on cumulative economic growth, unconditional of fiscal policy stance. The models are based on Equation 2. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics do not indicate cross-sectional dependence in any models. To account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.



**FIGURE 11** Average treatment effect (ATE) responses of Augmented Inverse Probability Weighted (AIPW) local projection estimates of the effect of product and labour market reforms occurring simultaneously on cumulative economic growth in  $t = 1, \dots, 5$ . *Note:* The figure plots the ATE responses of product and labour market reforms occurring simultaneously on cumulative economic growth based on regressions in Table 9 with the solid black lines. The y axis displays single-digit growth rates at the second decimal place. The dark grey shaded areas display the 90% error bands; the light grey shaded areas display the 95% error bands. Year  $t = 1$  is the first year after a reform took place at  $t = 0$

falsification test for product and labour market reforms unconditional on fiscal policy. Figures SA4.7–SA4.9 and SA4.10–SA4.12 report the results of the falsification test for product and labour market reforms conditional on fiscal policy. The simulated average treatment placebo effects are normally distributed around zero for all forecast horizons. Furthermore, the estimated significant effects are clearly placed in the tails of the distribution of the simulated average treatment placebo effects. This suggest that the significant ATEs that we find in our main analysis are not the result of type I errors.

In the case of labour market reforms, it is interesting to investigate whether there is a direct effect on employment as a precursor to seeing an effect on output. As shown in Tables 10 and 11 and Figure 12, the effect of labour market reforms on employment is largely consistent with our main results.

Next, as suggested by a reviewer, we have checked what happens if we use dummies that are only one in the first year of a sequence of new reforms in successive years. Table 12 shows the results. Not surprisingly, the biggest

**TABLE 10** Average treatment effect of labour market reforms on the change in employment unconditional of fiscal stance, AIPW estimates

	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Labour market reforms</b>	0.001 (0.001)	−0.001 (0.001)	−0.003* (0.002)	−0.003 (0.002)	−0.003 (0.002)
Pesaran test statistics	5.13	2.92	0.49	0.23	−0.20
Observations	545	522	499	476	453
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

*Note:* The table shows the ATE responses of the AIPW local projection estimates of labour reforms on the change in employment. The models are based on Equation 2. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.

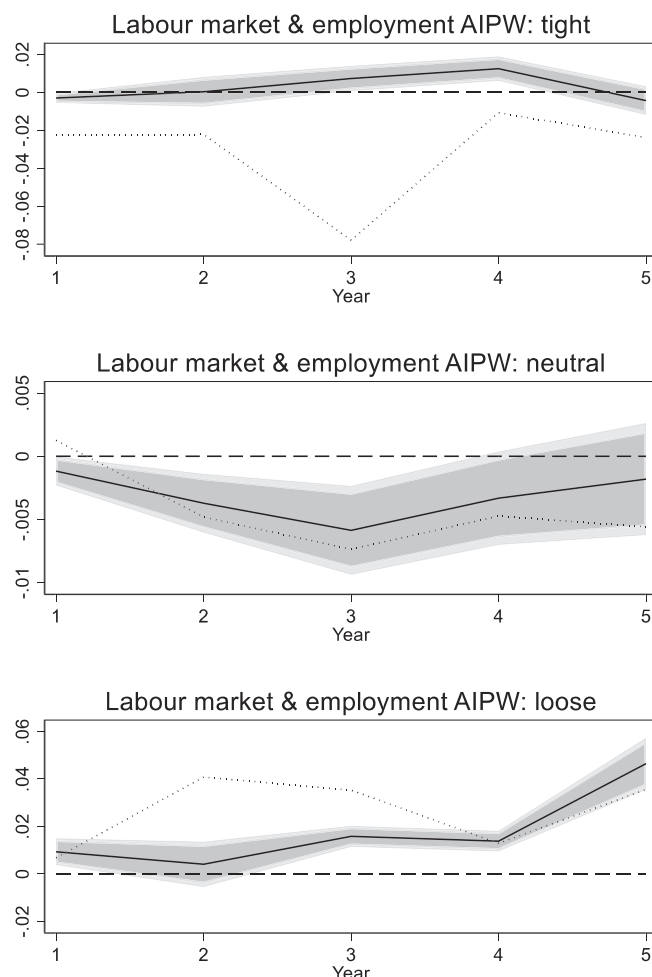
\* $p < 0.10$ .

**TABLE 11** Average treatment effect of labour market reforms on the change in employment conditional of fiscal stance, AIPW estimates

	(1) Year 1	(2) Year 2	(3) Year 3	(4) Year 4	(5) Year 5
<b>Labour market reforms</b>					
Tight	−0.003** (0.001)	0.000 (0.004)	0.007** (0.003)	0.013*** (0.003)	−0.004 (0.004)
Neutral	−0.001** (0.001)	−0.004*** (0.001)	−0.006*** (0.002)	−0.003* (0.002)	−0.002 (0.002)
Loose	0.009*** (0.003)	0.004 (0.005)	0.016*** (0.002)	0.014*** (0.002)	0.046*** (0.006)
Pesaran test statistic	1.01	1.42	4.34	1.97	−0.11
Observations	545	522	499	476	453
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

*Note:* The table shows the ATE responses of the AIPW local projection estimates of labour reforms on the change in employment, conditional of fiscal policy stance. Fiscal policy stance is identified using the structural break approach. The models are based on Equation 3. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.

\* $p < 0.10$ . \*\* $p < 0.05$ . \*\*\* $p < 0.01$ .



**FIGURE 12** Labour market reform and employment, conditional on fiscal policy. *Note:* The figure plots the ATE responses of labour market reforms on the change in employment based on the regressions reported in Table 10 with the solid black lines. The y axis displays single-digit employment changes in percentages of the population at the second decimal place. The dark grey shaded areas display the 90% error bands; the light grey shaded areas display the 95% error bands. Year  $t = 1$  is the first year after a reform took place at  $t = 0$ . For ease of comparison the dotted lines plot the AIPW responses for labour market reforms and growth

changes occur for product market reforms as there are more sequences with new product market reforms in successive years compared to new labour market reforms in successive years. Concerning labour market reform, the changes are minor.

Finally, a cause of concern about our estimates may be the Nickell (1981) bias. Specifically, we estimate a dynamic panel model with fixed-effects. As Nickell (1981) shows, the demeaning process creates a correlation between the regressor and the error term which creates a bias in the estimated coefficient of the lagged dependent variable. If the independent variables of interest are correlated with the lagged dependent variable their coefficients may be biased as well. This is particular a problem in a large  $N$ , small  $T$  context. We have small  $N$  and relatively large  $T$ . The bias can be gauged in the following way.

If the AR(1) coefficient  $\beta_2$  on  $\Delta y_{i,t}$  is positive (as in our case), the bias is invariably negative, so that the persistence of the  $\beta_2$  coefficient on  $\Delta y_{i,t}$  will be underestimated. For reasonably large values of  $T$ , the limit of  $\beta_2$  on  $\Delta y_{i,t}$  as  $N \rightarrow \infty$  will be approximately  $-(1 + \beta_2)/(T - 1)$ . In our case  $\beta_2 = 0.66$ , so that the bias will be about  $-0.062$ , that is, less than  $1/10$  of the estimated coefficient. This is even assuming that  $N$  tends to infinity, which is far from the case in our application. Furthermore, the correlation between the labour and product market indicators and  $\Delta y_{i,t}$  is low and negative. The correlation coefficient for product (labour) market reforms and lagged GDP growth is  $-0.04$  ( $-0.08$ ). Because of this negative correlation, the Nickell bias also leads to an underestimation of the impulse responses of reforms on

**TABLE 12** Average treatment effect of reforms dependent on the stance of fiscal policy when only the first reform in a reform sequences is considered, AIPW estimates

	Year 1	Year 2	Year 3	Year 4	Year 5
<b>Product market reforms</b>					
Tight	0.009** (0.004)	0.003 (0.003)	0.011*** (0.004)	0.009 (0.005)	0.005 (0.006)
Neutral	−0.001 (0.002)	−0.001 (0.002)	0.001 (0.002)	0.001 (0.003)	0.004 (0.003)
Loose	−0.004 (0.003)	−0.004* (0.003)	0.000 (0.003)	−0.014 (0.010)	−0.004 (0.015)
Pesaran test stat	5.28	1.75	5.67	−0.48	0.84
Observations	617	592	567	542	517
<b>Labour market reforms</b>					
Tight	0.001 (0.003)	0.001 (0.003)	0.007*** (0.002)	0.011*** (0.004)	0.007 (0.005)
Neutral	0.001 (0.001)	−0.004* (0.002)	−0.008** (0.003)	−0.005** (0.002)	−0.007* (0.004)
Loose	0.007* (0.003)	0.042** (0.019)	0.033*** (0.008)	0.011*** (0.004)	0.026*** (0.008)
Pesaran test stat	5.18	4.20	2.01	3.24	7.50
Observations	545	522	499	476	453
Treatment lags	5	5	5	5	5
Treatment leads = forecast horizon	Yes	Yes	Yes	Yes	Yes
Country fixed-effects	Yes	Yes	Yes	Yes	Yes

Note: The table shows the ATE responses of the AIPW local projection estimates of labour and product market reforms on cumulative economic growth, conditional of fiscal policy stance. Fiscal policy stance is identified using the structural break approach. Only the first reform in a reforms sequence is considered. The models are based on Equation 3. The number of treatment lags was determined by Akaike's information criterion. The Pesaran test statistics indicate cross-sectional dependence in some models, but to account for the imported uncertainty from the first-stage logit estimation we report cluster-bootstrapped standard errors with 500 repetitions in parentheses.

\* $p < 0.10$ . \*\* $p < 0.05$ . \*\*\* $p < 0.01$ .

growth. This, in combination with the relative low size of the biased AR(1) term and the large T relative to N leads us to conclude that the Nickell bias in our case is negligible.<sup>15</sup>

## 6 | CONCLUSIONS

Our findings indicate that controlling for endogeneity of reforms and the stance of fiscal policy is crucial. Our results suggest that product market reforms mostly cause slight negative growth. Labour market reforms hurt growth under tight and neutral fiscal policy but are conducive to economic growth if introduced during periods of expansionary fiscal policy.

One important topic for future research is to analyse the election effects of reforms. Recently, Alesina et al. (2020) found that liberalizing reforms are costly to incumbents when implemented close to elections. They also find that the electoral effects depend on the state of the economy at the time of reform: Reforms are penalized during contractions; liberalizing reforms undertaken in expansions are often rewarded. Our results suggest that in analysing the electoral consequences of reform, it is important to distinguish between labour and product market reforms, as they may affect economic growth differently, and to take the fiscal policy stance into account as well, since expansionary fiscal policy may alleviate the negative short-run growth effects of reform.

Another topic for future research is the spill-over effects of reform to other countries. We ignored this issue, while Lastauskas and Stakenas (2018) suggest that spill-overs may be important. Finally, the impact of reform on income inequality deserves further study.

<sup>15</sup>GMM estimation is not suited in cases of large T and small N. Rather a method based on recursive substitutions could be used. But as noted in Teulings and Zubanov (2014), a disadvantage of such an approach is a sizeable efficiency loss.

## ACKNOWLEDGMENTS

We like to thank two reviewers and participants in the 2020 conference of the European Economic Association for very useful feedback. Likewise, we thank Paul Elhorst, Robert Inklaar, Richard Jong-A-Pin, Milena Nikolova and Thanasis Ziogas for valuable help and feedback. We did not receive any funding for this project.

## REFERENCES

- Aghion, P., & Griffith, R. (2005). *Competition and growth: reconciling theory and evidence*. MIT Press.
- Alesina, A., & Ardagna, S. (2010). Large changes in fiscal policy: Taxes versus spending. In J. R. Brown (Ed.), *Tax policy and the economy* (Vol. 24, pp. 35–68). National Bureau of Economic Research, Inc.. <https://doi.org/10.1086/649828>
- Alesina, A., & Drazen, A. (1991). Why are stabilizations delayed? *American Economic Review*, 81, 1170–1188.
- Alesina, A., Favero, C., & Giavazzi, F. (2019). *Austerity: when it works and when it doesn't*. Princeton University Press.
- Alesina, A., Furceri, D., Ostry, J.D., Papageorgiou, C., & Quinn, D.P. (2020). Structural reforms and elections. Evidence from a world-wide new dataset. NBER Working Paper 26720. [https://www.nber.org/system/files/working\\_papers/w26720/w26720.pdf](https://www.nber.org/system/files/working_papers/w26720/w26720.pdf)
- Alesina, A., & Perotti, R. (1995). Fiscal expansions and adjustments in OECD countries. *Economic Policy*, 10, 205–248. <https://doi.org/10.2307/1344590>
- Auerbach, A. J., & Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4, 1–27. <https://doi.org/10.1257/pol.4.2.1>
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47–78. <https://doi.org/10.2307/2998540>
- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1–22. <https://doi.org/10.1002/jae.659>
- Banerji, A., Dabla-Norris, E., Duval, R., & Furceri, D. (2017). The case for fiscal policy to support structural reforms. <https://blogs.imf.org/2017/03/13/the-case-for-fiscal-policy-to-support-structural-reforms/>
- Bordon, A., Ebeke, C., & Shirono, K. (2018). When do structural reforms work? On the role of the business cycle and macroeconomic policies. In J. de Haan & J. Parlevliet (Eds.), *Structural reforms - moving the economy forward* (pp. 147–171). Springer International Publishing. [https://doi.org/10.1007/978-3-319-74400-1\\_7](https://doi.org/10.1007/978-3-319-74400-1_7)
- Borjas, G. J. (2017). The wage impact of the Marielitos: A reappraisal. *ILR Review*, 70, 1077–1110. <https://doi.org/10.1177/0019793917692945>
- Bouis R., Causa, O., Demmou, L., Duval, R., & Zdzienicka, A. (2012). The short-term effects of structural reforms: an empirical analysis. OECD Economics Department Working Paper 949. [https://www.oecd-ilibrary.org/economics/the-short-term-effects-of-structural-reforms\\_5k9csvk4d56d-en](https://www.oecd-ilibrary.org/economics/the-short-term-effects-of-structural-reforms_5k9csvk4d56d-en)
- Bouis, R., Duval, R., & Eugster, J. (2016). Product market deregulation and growth: new country-industry-level evidence. IMF Working Paper 16/114. <https://www.imf.org/external/pubs/ft/wp/2016/wp16114.pdf>
- Brueckner, M. (2021). Fiscal consolidations. CAMA Working Paper 57/2021. [https://cama.crawford.anu.edu.au/sites/default/files/publication/cama\\_crawford\\_anu\\_edu\\_au/2021-07/57\\_2021\\_brueckner0.pdf](https://cama.crawford.anu.edu.au/sites/default/files/publication/cama_crawford_anu_edu_au/2021-07/57_2021_brueckner0.pdf)
- Cacciatore, M., & Fiori, G. (2016). The macroeconomic effects of goods and labor market deregulation. *Review of Economic Dynamics*, 20, 1–24. <https://doi.org/10.1016/j.red.2015.10.002>
- Campos, N., & Nugent, J. B. (2018). The dynamics of the regulation of labor in developing and developed countries since 1960. In N. Campos, P. de Grauwe, & Y. Ji (Eds.), *The political economy of structural reforms in Europe* (pp. 75–88). Oxford University Press. <https://doi.org/10.1093/oso/9780198821878.003.0003>
- Carter, D. B., & Signorino, C. S. (2010). Back to the future: Modeling time dependence in binary data. *Political Analysis*, 18, 271–292. <https://doi.org/10.1093/pan/mpq013>
- Cette, G., Lopez, J., & Mairesse, J. (2016). Market regulations, prices, and productivity. *American Economic Review: Papers & Proceedings*, 106, 104–108. <https://doi.org/10.1257/aer.p20161025>
- Cole, S. R., & Hernan, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 186, 383–386. <https://doi.org/10.1093/aje/kwn164>
- Devries, P., Guajardo, J., Leigh, D., & Pescatori, A. (2011). A new action-based dataset of fiscal consolidation. IMF Working Paper 11/128. <https://www.imf.org/external/pubs/ft/wp/2011/wp11128.pdf>
- Draghi M. (2015). Structural reforms, inflation and monetary policy. Introductory speech by Mario Draghi, President of the ECB, ECB Forum on Central Banking Sintra, 22 May 2015. <https://www.ecb.europa.eu/press/key/date/2015/html/sp150522.en.html>
- Drazen, A., & Grilli, V. (1993). The benefit of crises for economic reforms. *American Economic Review*, 83, 598–607.
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80, 49–60. <https://doi.org/10.1162/003465398557825>
- Duval, R., & Furceri, D. (2018). The effects of labor and product market reforms: The role of macroeconomic conditions and policies. *IMF Economic Review*, 66, 31–69. <https://doi.org/10.1057/s41308-017-0045-1>
- Duval, R., Furceri, D., Hu, B., Jalles, J., & Nguyen, H. (2018). A narrative database of major labor and product market reforms in advanced economies. IMF Working Paper 18/19. <https://www.imf.org/en/Publications/WP/Issues/2018/01/25/A-Narrative-Database-of-Major-Labor-and-Product-Market-Reforms-in-Advanced-Economies-45585>



- Duval, R., Furceri, D., & Miethe, J. (2021). Robust political economy correlates of major product and labor market reforms in advanced economies: Evidence from BAMLE for Logit Models. *Journal of Applied Econometrics*, 36, 98–124. <https://doi.org/10.1002/jae.2791>
- Elhorst, J. P. (2013). Spatial panel models. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of regional science* (pp. 1637–1652). Springer International Publishing. [https://doi.org/10.1007/978-3-642-23430-9\\_86](https://doi.org/10.1007/978-3-642-23430-9_86)
- Faccini, R. (2014). Reassessing labour market reforms: Temporary contracts as a screening device. *The Economic Journal*, 124, 167–200. <https://doi.org/10.1111/ecoj.12072>
- Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105, 3150–3182. <https://doi.org/10.1257/aer.20130954>
- Fiori, G., Nicoletti, G., Scarpetta, S., & Schiantarelli, F. (2012). Employment effects of product and labour market reforms: Are there synergies? *The Economic Journal*, 122, F79–F104. <https://doi.org/10.1111/j.1468-0297.2011.02494.x>
- Glynn, A. N., & Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, 18, 36–56. <https://doi.org/10.1093/pan/mpp036>
- Guajardo, J., Leigh, D., & Pescatori, A. (2014). Expansionary austerity? International evidence. *Journal of the European Economic Association*, 12, 949–968. <https://doi.org/10.1111/jeea.12083>
- Haggard, S., & Webb, S. (1993). What do we know about the political economy of economic policy reform? *The World Bank Research Observer*, 8, 143–168. <https://doi.org/10.1093/wbro/8.2.143>
- Hibbs, D. A. (1977). Political parties and macroeconomic policy. *The American Political Science Review*, 71, 1467–1487. <https://doi.org/10.1017/S0003055400269712>
- Holden, S., & Larsson Midthjell, N. (2013). Successful fiscal adjustments. Does choice of fiscal instrument matter? CESifo Working Paper 4456. [https://www.econstor.eu/bitstream/10419/89663/1/cesifo\\_wp4456.pdf](https://www.econstor.eu/bitstream/10419/89663/1/cesifo_wp4456.pdf)
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86, 4–29. <https://doi.org/10.1162/003465304323023651>
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47, 5–86. <https://doi.org/10.1257/jel.47.1.5>
- IMF. (2016). *World Economic Outlook, Chapter 3*. International Monetary Fund.
- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95, 161–182. <https://doi.org/10.1257/0002828053828518>
- Jordà, Ò., & Taylor, A. (2015). The time for austerity: Estimating the average treatment effect of fiscal policy. *The Economic Journal*, 126, 219–255. <https://doi.org/10.1111/ecoj.12332>
- Lastauskas, P., & Stakenas, J. (2018). Structural labour market reforms in the EU-15: Single-country vs. coordinated counterfactuals. *Structural Change and Economic Dynamics*, 44, 88–89. <https://doi.org/10.1016/j.strueco.2017.11.001>
- Lütkepohl, H., & Kratzig, M. (2004). *Applied time series econometrics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511606885>
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107, 407–437. <https://doi.org/10.2307/2118477>
- Mierau, J. O., Jong-A-Pin, R., & de Haan, J. (2007). Do political variables affect fiscal policy adjustment decisions? New empirical evidence. *Public Choice*, 133, 297–319. <https://doi.org/10.1007/s11127-007-9189-2>
- Mink, M., Jacobs, J., & de Haan, J. (2012). Measuring coherence of output gaps with an application to the euro area. *Oxford Economic Papers*, 64, 217–234. <https://doi.org/10.1093/oeq/gpr049>
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49, 1399–1416. <https://doi.org/10.2307/1911408>
- Nicoletti, G., & Scarpetta, S. (2003). Regulation, productivity and growth: OECD evidence. *Economic Policy*, 36, 11–72. <https://doi.org/10.1111/1468-0327.00102>
- Parlevliet, J., Savsek, S., & Tóth, M. (2018). The impact of structural reforms: A review of the literature. In J. de Haan & J. Parlevliet (Eds.), *Structural reforms - moving the economy forward* (pp. 21–37). Springer International Publishing. [https://doi.org/10.1007/978-3-319-74400-1\\_2](https://doi.org/10.1007/978-3-319-74400-1_2)
- Perotti, R. (2012). The austerity myth: Gain without pain? In A. Alesina & F. Giavazzi (Eds.), *Fiscal policy after the financial crisis* (pp. 307–354). National Bureau of Economic Research, Inc.. <https://www.nber.org/system/files/chapters/c12652/c12652.pdf>
- Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34, 1089–1117. <https://doi.org/10.1080/07474938.2014.956623>
- Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, 126, 850–901. <https://doi.org/10.1086/696277>
- Syversen, C. (2011). What determines productivity? *Journal of Economic Literature*, 49, 326–365. <https://doi.org/10.1257/jel.49.2.326>
- Teulings, C., & Zubanov, N. (2014). Is economic recovery a myth? Robust estimation of impulse responses. *Journal of Applied Econometrics*, 29, 497–514. <https://doi.org/10.1002/jae.2333>
- Turrini, A., Koltay, G., Pierini, F., Goffard, C., & Kiss, A. (2015). A decade of labour market reforms in the EU: Insights from the LABREF database. *IZA Journal of Labor Policy*, 4, 1–33. <https://doi.org/10.1186/s40173-015-0038-5>
- Wiese, R., Jong-A-Pin, R., & de Haan, J. (2018). Can successful fiscal adjustments only be achieved by spending cuts? *European Journal of Political Economy*, 54, 145–166. <https://doi.org/10.1016/j.ejpoleco.2018.01.003>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** de Haan, J., & Wiese, R. (2022). The impact of product and labour market reform on growth: Evidence for OECD countries based on local projections. *Journal of Applied Econometrics*, 37(4), 746–770. <https://doi.org/10.1002/jae.2890>