INTERNATIONAL MONETARY FUND

Structural Reforms and Economic Growth: A Machine Learning Approach

Anil Ari, Gabor Pula and Liyang Sun

WP/22/184

2022
SEP

WORKING PAPER

 WP/22/184

IMF Working Paper
European Department

Structural Reforms and Economic Growth: A Machine Learning Approach
Prepared by Anil Ari, Gabor Pula and Liyang Sun

Authorized for distribution by Ivanna Vladkova Hollar
September 2022

ABSTRACT: The qualitative and granular nature of most structural indicators and the variety in data sources poses difficulties for consistent cross-country assessments and empirical analysis. We overcome these issues by using a machine learning approach (the partial least squares method) to combine a broad set of cross-country structural indicators into a small number of synthetic scores which correspond to key structural areas, and which are suitable for consistent quantitative comparisons across countries and time. With this newly constructed dataset of synthetic structural scores in 126 countries between 2000-2019, we establish stylized facts about structural gaps and reforms, and analyze the impact of reforms targeting different structural areas on economic growth. Our findings suggest that structural reforms in the area of product, labor and financial markets as well as the legal system have a significant impact on economic growth in a 5-year horizon, with one standard deviation improvement in one of these reform areas raising cumulative 5-year growth by 2 to 6 percent. We also find synergies between different structural areas, in particular between product and labor market reforms.

RECOMMENDED CITATION: Ari, A., Pula, G., & Sun, L. (2022). Structural Reforms and Economic Growth: A Machine Learning Approach. IMF Working Paper, WP/22/184

JEL Classification Numbers:	E02, C54, C55, D58, O43, O47
Keywords:	Structural reforms, institutions, economic growth
Author’s E-Mail Address:	aari@imf.org, gpula@imf.org, lsun20@cemfi.es

WORKING PAPERS

Structural Reforms and Economic Growth: A Machine Learning Approach

Prepared by Anil Ari, Gabor Pula and Liyang Sun[1]

[1] The authors are grateful to Ivanna Vladkova Hollar, Ippei Shibata, Marina Mendes Tavares and seminars participants at the IMF for helpful comments and suggestions. Excellent research assistance was provided by Samuel Victor Romero Martinez. The dataset of synthetic structural scores is available upon request from the authors. All errors are our own.

Contents

I. Introduction

Policymakers often pursue structural reforms to aid recovery from crises and stimulate economic growth. This places the onus on policymakers to identify which combinations and sequences of structural reforms would be the most growth-enhancing (IMF, 2015; Rodrik, 2010). However, a key challenge is that structural reforms are inherently difficult to measure as they often involve policies that are geared towards improving efficiency of markets. Common approaches quantify structural reforms based on the strength of regulatory changes that remove inefficiencies (see e.g., Alesina et al., 2020). While these approaches provide valuable insights on the impact of policy actions, they may not fully reflect reform outcomes, which depend on the specifics of policy implementation as well as the environment in which reforms are implemented. Another drawback of these approaches is that they have limited country coverage due to limited data availability. Other approaches rely on survey-based indicators of structural outcomes to assess the impact of structural reforms and conduct cross-country analysis (see e.g., Egert and Gal, 2016; Egert, 2017). While these indicators are informative about structural performance, empirical analysis is complicated by the large number of indicators, the correlation between them and biases that may arise from their subjective nature.

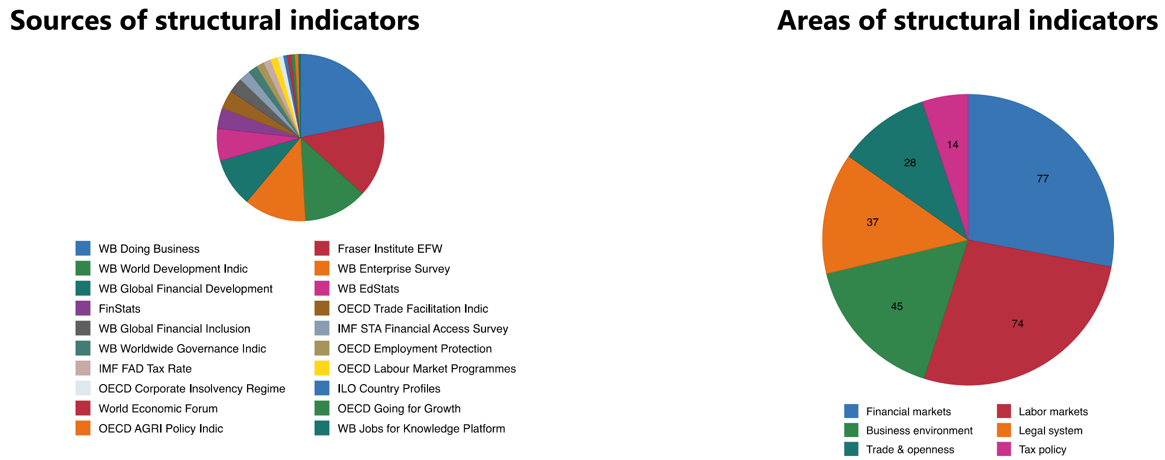
We use a machine learning approach to construct synthetic structural scores from a large number of structural indicators. Our analysis contributes to the existing literature by using partial least squares (PLS) to aggregate structural indicators for growth analysis, instead of simple averaging or ad hoc weighing schemes. Our PLS weighting scheme assigns higher weights to indicators that are more predictive of high GDP per capita, thereby extracting useful information from available data while removing the noise and biases associated with subjective and survey-based indicators. Our approach also accounts for the correlation and redundancy among structural indicators, therefore avoiding the duplication bias that simple averaging would suffer from.[1]

Our synthetic structural scores are based on a rich and disaggregated dataset of structural indicators. We rely on the IMF's Structural and Financial Indicators database which draws from several sources and includes 275 structural indicators from 126 countries (Figure 1).[2] We then group these indicators into six structural areas identified in IMF (2015): financial system (77), trade and openness (28), legal system (37), labor markets (74), business environment (45) and tax policy (14). We then construct a synthetic structural score for each structural area as the PLS-weighted average of the underlying structural indicators.

[1] Our approach builds upon Ari and Pula (2021) which proposes the use of principal component analysis (PCA), to form synthetic structural factors. The PCA weights account for the correlation between individual indicators but are sensitive to duplication of indicators, which is common in our dataset due to overlaps in data sources.

[2] Our analysis includes indicators from the World Bank's Doing Business (DB) dataset, which has recently been suspended due to concerns about data manipulation. While this poses a drawback for our study as well as a significant portion of the literature on structural reforms, it is worth noting that this is the form of subjectivity bias that we aim to alleviate with our PLS approach.

Figure 1. Structural indicator overview



Using the synthetic structural scores, we find significant growth impacts from reforms in certain structural areas, as well as synergies between different structural areas. Our findings suggest that structural reforms in the areas of product, labor and financial markets as well as the legal system have a significant impact on economic growth in a 5-year horizon, with one standard deviation improvement in one of these reform areas raising cumulative 5-year growth by 2 to 6 percent. We also find synergies between different structural areas, in particular between product and labor market reforms.

The paper is organized as follows. Section II overviews the data and discusses our approach to imputing missing indicators. Section III applies PLS to construct synthetic structural scores based on the imputed indicators, controlling for the correlation among the individual structural indicators, and assigning the weights to reflect how predictive the indicators are for output. Section IV uses the synthetic structural scores to analyze the impact of structural reforms on growth. Finally, Section V concludes.

II. Structural Indicators

A. Data overview

The performance of structural reforms is measured using quantitative indicators. Cross-country data on a large set of structural indicators are obtained from the Fund's Macrostructural Database, which combines data from several sources. These indicators are then categorized to six broader macrostructural areas, listed as:

- *Legal system,* which includes structural indicators related to corruption, governance, crime, the rule of law and the protection of property rights.

- *Financial system,* which covers structural indicators pertaining to financial development, access to financial services and the soundness of the banking sector and financial markets.

- *Product markets,* which contains structural indicators on competition, informality, and administrative and regulatory burdens in product markets.

- *Labor markets,* which includes structural indicators related to minimum wages and other regulations that affect labor market flexibility.

- *Tax policy,* which captures distortions in incentives associated with various taxes.

- *Trade and openness,* which covers tariffs and non-tariff barriers to trade.

We exclude cyclical financial indicators, which reflect the business cycle rather than quality of financial institutions.[1]

Data coverage varies a lot by country and year, and the missing pattern is systematic as opposed to missing-at-random. For example, several indicators are only updated every other year while coverage for several indicators only start in recent years. As a first pass in imputing missing values, we take five-year averages of indicators starting in the year of 2000. To avoid deflating the variance, we only retain the data for every five years. We then exclude indicators that missing more than 20% of the values and impute the rest of missing values by a multiple imputation procedure as described in Appendix 0 There is no simple recommendation for a maximum proportion of missing values that can be properly considered in imputation methods. The results start to be unstable for a threshold above 20% and we leave it to future research to gauge the optimal amount of imputation.

[1] Examples of the cyclical financial indicators are the volume of total syndicated loans issued and availability of private credit.

B. Synthetic structural scores via Partial Least Squares

Given the high number of structural indicators, dimensionality reduction is necessary to improve interpretability for further analysis. We make the following observations on these indicators:

- We want to capture indicators associated with strong economic performance, which can be measured with the ability to predict high future per capita GDP.
- The indicators can be highly correlated within and across structural areas.
- We have many indicators relative to the sample size. There are 275 indicators, which is substantial compared to a sample size of 504 (126 countries and 4 time periods in 2000-04, 2005-09 and 2010-14, 2015-19).

These observations motivate the appropriate approach to dimensionality reduction. The naïve approach for prediction is to estimate a linear regression on these indicators, and use the predicted value as the composite score. However, when there are many correlated variables in a linear regression model, their coefficients can become unstable: a large positive coefficient on one indicator can be canceled by a similarly large negative coefficient on its correlated indicator. LASSO improves upon linear regression in allowing for high-dimensional indicators, which assumes there are only a few predictors for the outcome variable. While this assumption is more likely to hold in certain settings such as predicting non-performing loans in Ari et al. (2021), it is unlikely to hold for our outcome variable, log of future per capita GDP (in PPP). Consequently, LASSO would reduce the dimension too much and result in poor predictive performance.

Another common dimensionality reduction technique is principal component analysis (PCA), which seeks a weighted average of the indicators that have high variation across countries. This has the advantage of making full use of the available information to minimize noise related to any individual structural indicator and it also provides a weighting scheme that accounts for the correlation between individual indicators. However, this approach performs poorly when we have redundant indicators.[2]

Partial least squares (PLS) is a flexible machine learning technique that achieves both goals and is appropriate for our setting (Hastie et al., 2009). PLS improves upon PCA by adding a predictive model. To receive high weights under the PLS weighting scheme, the indicators also need to be predictive of the outcome. PLS also improves linear regression by accounting for the correlation between individual indicators. Unlike LASSO, PLS does not assume only few indicators are predictive of the outcome. In Appendix C, we also illustrate the advantages of PLS compared to scores that are based on simple averages of structural indicators. Below we provide further details about the PLS method.

[2] For example, the economic freedom index from the Fraser Institute is constructed based on data from WDI, the WEF GCR, the WGI, and the WB Doing Business. Therefore the Fraser Institute indicators are redundant when we include their source indicators. While it is possible to manually remove such redundancy based on a careful examination of the data sources of the indicators, we focus on a data-driven approach.

C. PLS estimation procedure

Let $x_{i,t}^{c}$ denote the vector of the indicators in country i at time t. Each indicator vector is from one of the six structural areas c. Let the index j further denote the subcategory of the indicator within the structural area. The PLS method estimates the following predictive model for the five-year-ahead per capita GDP ($y_{i,t}$):

$$y_{i,t} = \alpha + \sum_{c} \sum_{m} \theta_m \sum_{j} \gamma_{j,m}^{c}\, x_{i,t,j}^{c} \; + \epsilon_{i,t} \tag{1}$$

where m indexes the number of components used. Because the LHS of Equation (1) is the five-year-ahead per capita GDP, we use the largest possible sample with indicators from 2000-2010 to estimate Equation (1). However since we aim for good predictive performance, we cannot just choose the number of components to maximize the in-sample fit for 2000-2010 when we estimate Equation (1). We therefore use leave-one-out cross-validation to determine the number of components, which suggests that eight components provide the best predictive performance.

Unlike the linear method that minimizes the in-sample prediction error, the PLS method estimates equation (1) using an iterative procedure consisting of the four steps described below. This procedure provides an implicit regularization on the magnitude of the coefficients (see step 4) of the procedure, which improves upon the linear method. Operationally, we use the R library plsr to implement the PLS.

Initialize the right-hand-side (RHS) of equation (1) with the original data $x_{i,t}^{c,(0)} = x_{i,t}^{c}$ and initialize the predicted left-hand-side (LHS) with the sample mean of the outcome $\hat{y}_{i,t}^{(0)} = \bar{y}$. Note the RHS is standardized to be mean zero and standard deviation one. For the m-th components, the PLS algorithm proceeds as the follows:

1) Form the component based on the original inputs

$$z_m = \sum_{c} \hat{\gamma}_m^{c} x_{i,t}^{c,(m-1)} \tag{2}$$

   where $\hat{\gamma}_m^{c} = cov(x_{i,t}^{c,(m-1)}, y_{i,t})$ is the covariance between the original inputs and the outcome;

2) Calculate the coefficient in front to the component as the OLS coefficient of regressing the outcome on the component

$$\hat{\theta}_m = \frac{Cov(y_{i,t}, z_m)}{Var(z_m)} \tag{3}$$

3) Predict the outcome using all components so far as

$$\hat{y}_{i,t}^{(m)} = \hat{y}_{i,t}^{(m-1)} + \hat{\theta}_m z_m \tag{4}$$

4) Orthogonalize $x_{i,t}^{c,(m-1)}$ with respect to the component z_m to get the updated input $x_{i,t}^{c,(m)}$. This ensures the next component z_{m+1}, which is a weighted average of $x_{i,t}^{c,(m)}$, is uncorrelated with z_m. The correlation across indicators is accounted for in this step. Furthermore, the updated input $x_{i,t}^{c,(m)}$ is a weighted average of the original inputs $x_{i,t}^{c,(m-1)}$, with weights reflecting the covariance across the original inputs and their covariance with the outcome.

D. Synthetic structural score as the predicted value from the PLS model

We construct the synthetic structure score for 2000-2015 in a given category c as the predicted value from the PLS model (1), predicted using the 2000-2015 indicators. Specifically, the synthetic structural score is the predicted value using the PLS coefficient estimates $\hat{\theta}_m$ and all components z_m:

$$\sum_m \hat{\theta}_m z_m = \sum_m \hat{\theta}_m \sum_c \hat{\gamma}_m^c x_{i,t}^{c,(m-1)} \tag{5}$$

which is a weighted average of the original indicators as explained above. Therefore, we can examine the indicators that receive the largest weights to confirm whether the composite scores are interpretable.

Table 1 tabulates indicators with large weights for each structural area. These subcategories mostly coincide with those selected by IMF (2019) to assess structural performance in these areas. This provides credibility to the PLS method for selecting highly interpretable subcategories in constructing the scores. To make scores comparable across structural areas, for further analysis we standardize each score to have zero mean and unit variance. If a country scores one in the financial composite but minus one in the business environment composite, then this can be interpreted as its structural performance in the financial area contributing one standard deviation more to its per capita GDP than an average country, while the opposite is true for its performance in the business environment structural area.

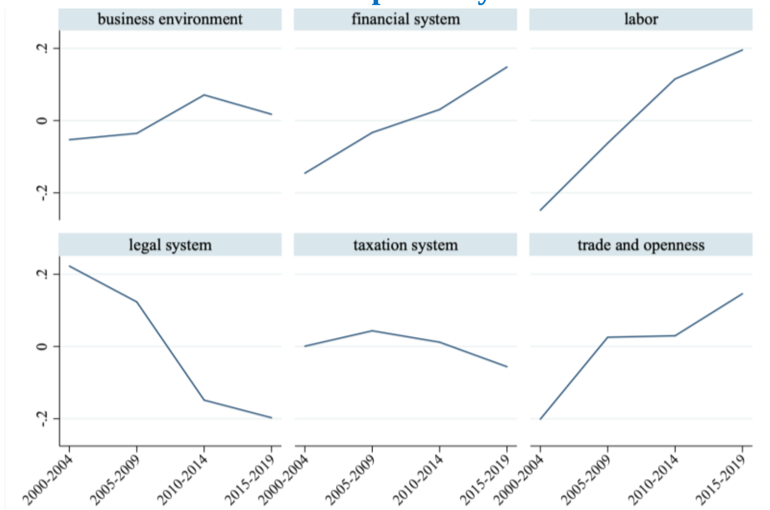
Table 1. Key structural indicators in the composition of composite structure scores

Business environment	Financial system	Labor market	Legal system	Tax policy	Trade and openness
Business infrastructure	Financial access	Education	Corruption	Tax rate	Tariffs
Mobile phone subscriptions (per 100 people); Access to electricity (time, cost etc.)	Account at financial institution; Number of bank branches per population etc.	Gross enrollment ratios,	Bribery index, % of firms expected to give gifts during public transactions	Top marginal tax rate, top marginal income tax, labor tax and contributions	Revenue from trade taxes; tariff rate
Starting business	Financial depth	Hiring / firing	Enforcement of contracts	Tax administration	Non-tariff barriers
Time, cost, number of procedures	Financial system deposits / GDP; Stocks traded value / GDP	Notice period, severance pay for redundancy dismissal	Time, cost, number of procedures	Ratio of VAT gap to VTTL, C-Efficiency ratio	Compliance cost of exporting and importing; trade facilitation performance
Admin and regulatory burden	Financial efficiency	Wage flexibility	Rule of law		Capital controls
Time spent with complying with regulations; Government efficiency	ROA, ROE, Net interest margin	Minimum wage, hours regulation, collective bargaining	Protection of property rights; judicial independence, integrity		Controls to capital movement, black market exchange rate
Informality	Financial stability	Inclusiveness			
Firms formally registered	Liquidity ratio, CAR etc.	Equal wage, no discrimination at hiring			

Note: Grey lines represent subgroups of indicators that contribute heavily toward the composite scores within each structural area. Each line below the grey line lists examples of the structural indicator in these subgroups.

Based on the composite indicators, there are flattening trends across the structural areas as shown in Figure 2. Since structural composites are constructed to predict output levels, an upward trend in the composite can be interpreted as a measure for structural reforms (i.e., improvements in structural performance), and the slopes of these trends can be interpreted as a measure for reform speed. There have been structural reforms in most areas except legal system, where reforms slowed down in recent years. This trend is consistent with IMF (2019), which finds stabilization of policies in many structural areas in late 2000s, and no improvement in the legal system. Instead of structural indicators, IMF (2019) measures structural reforms based on deregulations. The fact that composite scores present similar stylized facts with IMF (2019) lends validity to our approach of aggregating structural indicators.

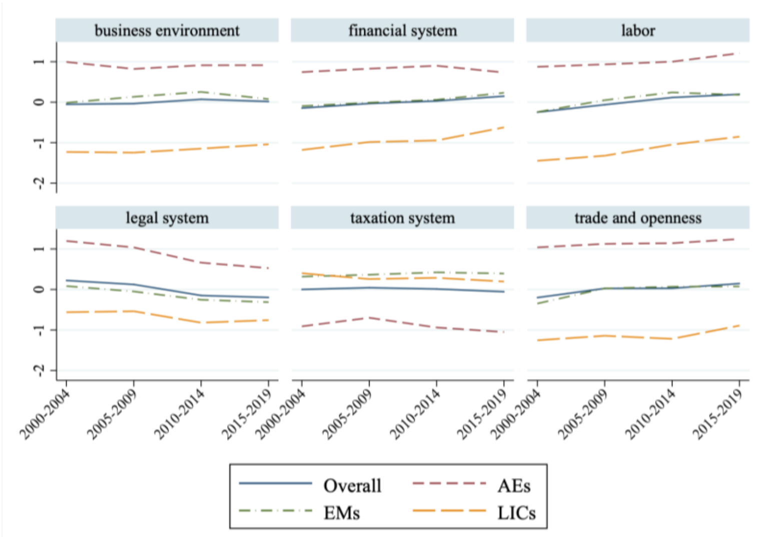
Figure 2. Trends of structural composite by structural areas



Note: The horizontal axis indicates the five-year window the structural indicators are collected. The lines plot the average of the composite across countries. The composites are standardized to have zero mean and unit variance across all countries and years.

The pattern of structural reform varies across income region as shown in Figure 33. There is a large gap between the structural composites of EMs and LICs and those of AEs in the area of business environment, labor market, legal system and trade and openness. Despite strong push for reforms, this gap suggests that EMs and LICs have substantial reform deficit in these areas, in line with the conclusion of IMF (2019). Nonetheless, LICs have shown improvement in labor market, reflected in an upward trend in the composite scores. Table 2 further shows the slowdown in legal reforms is common to all geographic regions.

Figure 3. Trends of structural composite across income groups



Note: The horizontal axis indicates the five-year window the structural indicators are collected. The lines plot the average of the composite across countries in a given region. The composites are standardized to have zero mean and unit variance across all countries and years.

Table 2. Share of countries that experience increases in structural composites

	Business environment			Financial system			Labor market			Legal system			Tax system			Trade and openness		
	2005-2009	2010-2014	2015-2019	2005-2009	2010-2014	2015-2019	2005-2009	2010-2014	2015-2019	2005-2009	2010-2014	2015-2019	2005-2009	2010-2014	2015-2019	2005-2009	2010-2014	2015-2019
Adv.Asia	80%	60%	60%	20%	20%	60%	80%	80%	60%	60%	0%	40%	80%	20%	0%	80%	80%	60%
EU	30%	53%	43%	60%	57%	40%	73%	47%	83%	17%	3%	7%	70%	37%	47%	63%	50%	77%
OtherAdv	50%	100%	50%	100%	50%	0%	100%	50%	50%	0%	0%	50%	50%	50%	0%	50%	50%	100%
Em.Asia	64%	71%	50%	50%	50%	57%	86%	86%	79%	50%	36%	57%	43%	64%	14%	79%	64%	50%
Em.EUR	70%	30%	50%	50%	80%	90%	90%	40%	90%	30%	40%	40%	70%	70%	20%	80%	70%	50%
MENA	56%	50%	44%	69%	44%	50%	88%	81%	38%	38%	31%	44%	44%	56%	25%	94%	56%	31%
SSA	32%	45%	77%	59%	27%	68%	82%	77%	82%	45%	32%	59%	27%	27%	32%	77%	41%	73%
LAC	75%	60%	20%	65%	55%	60%	100%	80%	20%	15%	15%	65%	40%	35%	40%	90%	35%	65%
OtherEM	33%	67%	50%	50%	50%	83%	67%	67%	100%	50%	33%	67%	67%	83%	33%	33%	33%	83%

Note: Red reflects low reform activities. Green reflects reform activities.

III. Structural Indicators

A. Impact of structural reforms on growth

The impact of structural reforms on growth is first estimated using cross-country regressions. Let i and t index country and each of year windows: 2000-2004, 2005-2009, 2010-2015, and 2016-2019. We take the five year average of GDP growth rate $g_{i,t}$. We use $S^s_{i,t}$ to denote the composite structural score for each of the six structural areas. We measure structural reform as the change in the structural composite, which is denoted with $\Delta S^s_{i,t-1}$. Since we are interested in marginal impact of structural reform in any given area, holding other areas constant, we estimate the 5-year cumulative growth impact using the following regression specification

$$g_{i,t} = a_i + \gamma_t + \sum_s \beta^s \, \Delta S^s_{i,t-1} + \sum_x \psi^x \, x_{i,t-1} + \epsilon_{i,t}$$

where a_i is a vector of country region fixed effects and dummies for emerging market economies and oil exporters,[3] and γ_t are time fixed effects. The set of control variables $x_{i,t}$ includes initial economic conditions as measured by the per capita GDP level in 2000, the VIX volatility index interacted with external debt, the VIX volatility index interacted with current account deficits, and vulnerability to oil price shocks as measured by the interaction of the oil exporter dummy with oil prices. The regression coefficient β^s can be interpreted as the reform elasticity of growth for a given structural area.

Table 3 presents the estimated regression coefficients, and each column varies the specification by alternating fixed effects and control variables. The estimates are robust to various specifications: one standard deviation increase in the composite scores for business environment, financial and labor markets, legal system have a significant positive impact on growth, ranging from 2 to 6%, holding other structural areas constant. However, trade and tax policy reforms have a statistically insignificant impact on growth.

[3] These are constructed using IMF WEO country classifications.

Table 3. Growth impact of structural reforms

VARIABLES	(1) GDP growth	(2) GDP growth	(3) GDP growth	(4) GDP growth
Business environment	0.0321**	0.0333**	0.0330***	0.0230**
	(0.0127)	(0.0130)	(0.00991)	(0.0111)
Financial markets	0.0196**	0.0193**	0.0182**	0.0189**
	(0.00869)	(0.00849)	(0.00799)	(0.00833)
Labor markets	0.0546***	0.0543***	0.0640***	0.0640***
	(0.0176)	(0.0172)	(0.0198)	(0.0190)
Legal system	0.0338**	0.0315**	0.0330***	0.0238**
	(0.0132)	(0.0132)	(0.0123)	(0.0109)
Tax policy	0.00120	0.000372	0.0125	0.00659
	(0.0115)	(0.0114)	(0.0111)	(0.00975)
Trade & openness	0.00945	0.00607	0.00641	0.00864
	(0.0139)	(0.0136)	(0.0123)	(0.0127)
Observations	251	251	224	224
R-squared	0.094	0.099	0.363	0.487
Time FE	NO	YES	YES	YES
Country group FE	NO	NO	NO	YES
Controls	NO	NO	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

These estimates might suffer from endogeneity bias. A favorable macroeconomic environment might induce small structural reforms, and at the same time lead to economic growth. To minimize such bias, we limit attention to structural episodes as measured by large discrete jumps in structural composites, in an attempt to capture policy initiatives that are not driven by current macroeconomic conditions. This approach has been used by Reinhart and Rogoff (2011), which define "inflation crisis" based on threshold for inflation rate. Studies on the impact of an increase in public investments have taken a similar approach by defining investment booms based on threshold for public-investment-to-GDP ratio (see e.g. IMF 2014, Warner 2014, and Ari et al. 2020.) Specifically, we define a "reform episode" to be an improvement in the top 10 percentile of positive changes in composite scores $\Delta S_{i,t-1}^{s}$. To shed light on the importance of sustained structural reform efforts, we also identify "reversed reform episodes" where top 10 percentile positive changes in composite scores $\Delta S_{i,t-1}^{s}$ is followed by a top 10 percentile negative change. Table 4 tabulates the distribution of reform episodes by structural areas. By definition the number of reform episodes is evenly distributed across structural areas, and about one third of reform episodes are reversed.

Table 4. Distribution of reform episodes

Structural area	**No. of episodes**	**Reversed**	**2005-2009**	**2010-2014**	**2015-2019**
Business environment	38	11	12	17	9
Financial system	38	10	10	15	13
Labor markets	38	11	12	15	11
Legal system	38	8	15	10	13
Tax policy	37	5	15	16	6
Trade & openness	38	6	18	5	15
TOTAL	227	51	82	78	67

We estimate the 5-year cumulative growth impact of reform episodes using the following regression specification

$$g_{i,t} = a_i + \gamma_t + \sum_{s} \beta^s E_{i,t-1}^{s} + \sum_{s} \phi^s N_{i,t-1}^{s} + \sum_{x} \psi^x x_{i,t-1} + \epsilon_{i,t}$$

where $E_{i,t}^{s}$ is an indicator for reform episodes and $N_{i,t}^{s}$ is an indicator for non-reversed reform episodes. The regression coefficient β^{s} captures the growth impact of a reform episode, and ϕ^{s} captures the differential impact of sustained (i.e., non-reversed) reform. Table 5 shows while reform episodes in business environment and labor market still have significantly positive impact, reform episodes in other structural areas no longer have a statistically significant impact. Reform episodes in legal systems only have a significant impact when not reversed later.

Table 5. Growth impact of structural episodes

Structural area	Reform episodes	Non-reversed
Business environment	0.0718**	-0.0419
	(0.0299)	(0.0329)
Financial markets	0.0270	-0.0364
	(0.0342)	(0.0405)
Labor markets	0.0321**	0.00814
	(0.0153)	(0.0249)
Legal system	-0.0269	0.0524*
	(0.0259)	(0.0313)
Tax policy	0.0193	-0.0373
	(0.0374)	(0.0379)
Trade & openness	0.0251	-0.00225
	(0.0259)	(0.0338)
Observations	224	
R-squared	0.483	
Time FE	YES	
Country group FE	YES	
Controls	YES	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

These results complement the existing literature on the growth impact of structural reforms. We focus on papers that construct structural indicators for different structural areas, though their methods of aggregating structural reforms into structural indicators differ. Differences in time and country coverage also make the direct comparison difficult. Egert (2017) found that one standard deviation increase in government effectiveness in business is associated with 7.7% increase in five-year growth rate, similar to the large impact we identified. IMF (2019) identified medium-run growth impact of a major reform (that increase their structural indicator by two standard deviation) in different structural areas (Ch.3, p.102). Reforms in the financial markets led to only one to two percent increase in growth. Among advanced economies, Duval and Furceri (2018) found labor market reforms are associated with 0.3 to 1.8% five-year growth impact. Furceri, Loungani and Ostry (2017) found a small but negative impact from capital account liberalization, which is consistent with the insignificant impacts of reform in the structural area of trade & openness found in our analysis.

B. Synergies of structural reforms on growth

Interactions among reforms across structural areas are oftentimes complementary. For example, IMF (2019) found deregulations in the product market only made a large impact when governance was strong in the reform country. To analyze these interactions, we create a threshold dummy $D_{i,t}^{z}$ for whether the country is in the top tercile in terms of the composite $S_{i,t}^{z}$ in structural area z. We then interact this threshold dummy with the change in the composite $\Delta S_{i,t-1}^{s}$ of various areas. Consider the following regression specification

$$g_{i,t} = a_i + \gamma_t + \sum_{s} \beta^s \, \Delta S_{i,t-1}^{s} + \sum_{s} \phi^{zs} \, D_{i,t-1}^{z} \Delta S_{i,t-1}^{s} + \sum_{x} \psi^x \, x_{i,t-1} + \epsilon_{i,t}$$

The regression coefficient ϕ^{zs} measures the incremental growth impact of structural reforms in area s when structural performance in area z is strong. Table 6 presents the coefficient estimates, where each column corresponds to the regression where we include the interaction between the composite and the threshold dummy in a given area.

Three findings stood out as highlighted in estimates of boldface.

First, there is positive synergy between business environment and labor market. We estimate that reforms strengthening the business environment have a larger growth impact when the labor market has a strong structural performance and vice versa. Given that the labor market composite reflects indicators of human capital and labor market flexibility (see Table 1), these findings suggest that the benefits associated with a favorable business environment may only be unlocked when there is a well-educated labor force and a flexible labor market. These findings are also consistent with Duval and Furceri (2018), who find that when deregulation makes labor markets are more flexible, firms tends to increase their hiring more in response to a rise in demand for their products.

Second, reforms in business environment have a larger growth impact when the legal system is strong. The composite for the legal system consists mostly of property rights protection. The positive synergy of business environment is therefore in line with Acemoglu, Johnson and Robinson (2005)'s finding that returns to an improvement in business environment is larger when property rights are well enforced.[4]

Third, reforms in labor markets are not as effective where there is high degree of openness to trade and capital flows, while reforms in business environment are. With low barriers to trade, countries are better able to substitute domestic labor (e.g., via outsourcing), which may limit the gains from labor market reforms. On the other hand, business environment reforms would likely yield greater benefits when businesses face international competition and an open capital account that provides easier access to foreign direct investment (FDI).

[4] In Section IV of their paper, Acemoglu, Johnson and Robinson (2005) regressed urbanization on the potential for Atlantic trade, which can be considered as a measure for business environment, and institutional qualities across European countries. The measure for initial institutions is based on how absolutist each country is, and the authors argued that property rights emerged in less absolutist countries. The authors found a statistically significant additional return from the potential for Atlantic trade among countries with better institutional qualities. This provides historical evidence that returns to an improvement in business environment is larger when property rights are well enforced.

Table 6. Synergies between structural areas

	Threshold dummy structural area (Z)					
Structural area (S)	**Business env.**	**Financial markets**	**Labor markets**	**Legal system**	**Tax policy**	**Trade & openness**
Business environment		0.0327	**0.0491***	**0.0469****	0.00971	**0.0461***
		(0.0228)	**(0.0289)**	**(0.0209)**	(0.0204)	**(0.0268)**
Financial markets	-0.00659		-0.0180	-0.0141	-0.00930	-0.00821
	(0.0154)		(0.0156)	(0.0168)	(0.0166)	(0.0137)
Labor markets	**0.0579***	-0.0416		0.0247	0.0371	**-0.0619***
	(0.0305)	(0.0358)		(0.0450)	(0.0302)	**(0.0321)**
Legal system	-0.0301	-0.0474	-0.00329		-0.0104	-0.0208
	(0.0237)	(0.0314)	(0.0307)		(0.0209)	(0.0306)
Tax policy	0.0161	-0.0110	0.00330	-0.0101		-0.0262
	(0.0184)	(0.0221)	(0.0220)	(0.0228)		(0.0218)
Trade & openness	-0.0441	-0.0346	-0.0271	-0.00492	-0.00824	
	(0.0269)	(0.0337)	(0.0301)	(0.0455)	(0.0225)	
Observations	224	224	224	224	224	224
R-squared	0.500	0.506	0.501	0.498	0.494	0.504
Time FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Country group FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

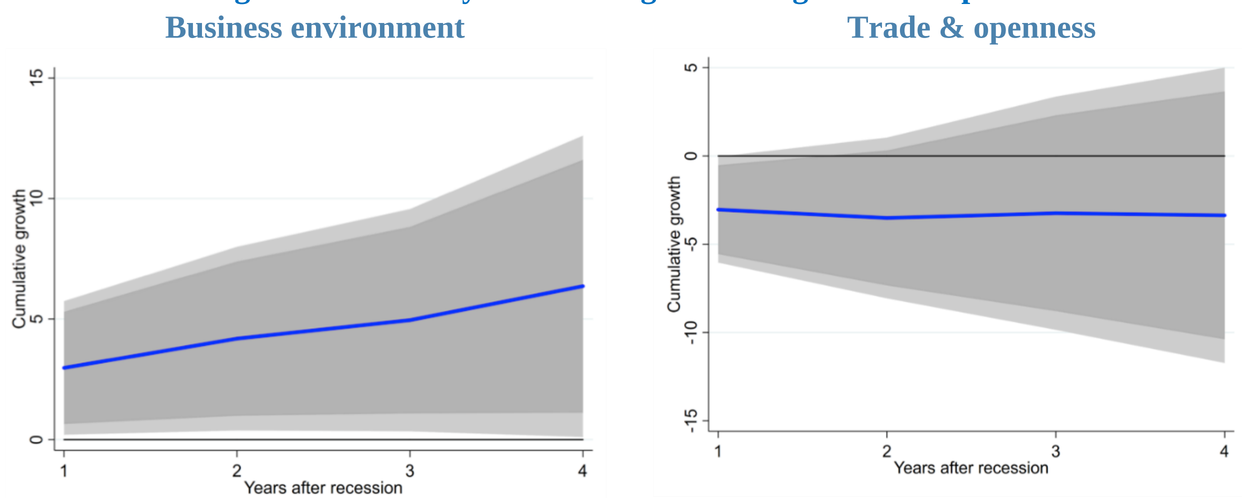
C. The role of structural reforms during crises

Countries sometimes implement structural reforms in response to economic crises. We identify 124 recessions in our dataset based on a negative annual growth rate and excluding repeated "recessions" that happen during a five-year interval (as the cumulative growth rate would be affected by the previous recession). We apply the local projection method (Jordà, 2005) to study the dynamic impact of pre-crisis structural performance on the growth rate after the recession. Specifically, for $h = 1,2,3,4$ years after the onset of the recession, we estimate the following specification for a given structural area s

$$g_{i,t+h} = a_i + \gamma_t^{GFC} + \beta_h^s D_{i,t-1}^s + \sum_x \psi_h^x x_{i,t-1} + \epsilon_{i,t+h}$$

where γ_t^{GFC} is an indicator for the global financial crisis (2007, 2008, 2009) and $D_{i,t}^s$ is an indicator for whether the country is in the top tercile in terms of the composite $S_{i,t}^s$ in structural area s. We control for pre-crisis per capita GDP and GDP growth as a measure for overheating, as well as the pre-crisis current account deficit and external debt.

Figure 4. Recovery from crisis given strong structural performance



Source: IMF staff calculations

Note: the lines denote the differential growth rate impact in percent between countries in the top tercile in terms of the structural composite versus the rest countries; the shaded areas denote 90 percent (darker) and 95 percent (lighter) confidence bands.

Error! Reference source not found.4 plots the estimates for β_h^s for each $h = 1,2,3,4$ years from the onset of the recession in blue lines, which are estimates for the differential growth impact for countries with strong structural performance over different horizons since the recession. A strong business environment is inducive to a speedy recovery, with a 3% increase in the year following the onset of the recession, and cumulatively a 5% increase after four years. These impacts are all statistically significant at the 5% level. We also find suggestive evidence that a high degree of trade and openness may reduce countries' resilience, with a 2.5% decrease in growth rate in the year following the onset of the recession, though cumulatively the impact is statistically insignificant. The impact from the other structural areas is not statistically distinguishable from zero.

IV. Conclusions

Structural reforms are difficult to quantify, as countries need to adapt measures to their specific settings. The conventional approach is to examine policy changes pertaining to relevant structural areas. A major impediment to applying this approach more broadly is the expertise required to understand policies in different countries. In this paper we consider a new approach that leverages a wide range of structural indicators compiled for a broad set of countries. We apply a machine learning approach, the partial least square method to aggregate high-dimensional structural indicators into composites for six structural categories. We then use these structural composites to estimate the growth impact of structural reforms, and find that the 5-year cumulative impact ranges from 2½ to 6½ percentage points depending on the structural area.

The structural composites also allow us to investigate synergies between structural areas, and whether strong structural performance helps countries recover from economic crises. We find that reforms in business environment lead to larger growth impact when labor market performs well, as measured by a high composite score. The reverse also holds, which suggests on business environment and labor market reforms are complementary. Focusing on countries that recovered from economic crises, we also find that strong pre-crisis performance in business environment helps with post-crisis recovery.

Findings of this paper make a strong case for use of machine learning in quantifying structural reforms. Machine learning provides a scalable approach to aggregate diverse sources of data. This paper also provides a template for how machine learning can be used to construct composites for finer structural areas. Growth analysis with finer composites could provide more empirical support for the Fund to determine reform priorities.

References

Acemoglu, Daron, Simon Johnson, and James Robinson. 2005. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review*, 95 (3): 546-579.

Alesina, A., Furceri, D., Ostry, J., Papageorgiou, C. and Quinn, D., (2020), Structural Reforms and Elections: Evidence from a World-Wide New Dataset, No 14371, CEPR Discussion Papers, C.E.P.R. Discussion Papers.

Ari, A., Lev., R and Chen, S. (2021). The dynamics of non-performing loans during banking crises: a new database with post-COVID-19 implications. *Journal of Banking and Finance*, forthcoming.

Ari, A., Bartolini, D., Boranova, V., DiBella, G., Dybczak, K., Hongjo, K., Huidrom, R., Jobst, A., Jovanovic, N., Ozturk, E., Papi, L., Sola, S., Stone, M., Topalova, P., "Infrastructure in Central, Eastern, and Southeastern Europe: Benchmarking, Macroeconomic Impact, and Policy Issues," IMF Departmental Papers / Policy Papers 2020/011, International Monetary Fund.

Ari, A. and Pula, G. (2021). Assessing the Macroeconomic Impact of Structural Reforms in Ukraine, IMF Working Papers, 2021(100)

Romain Duval & Davide Furceri, 2018. "The Effects of Labor and Product Market Reforms: The Role of Macroeconomic Conditions and Policies," IMF Economic Review, Palgrave Macmillan;International Monetary Fund, vol. 66(1), pages 31-69, March.

Egert, B., and Gal, P. (2016). The quantification of structural reforms in OECD countries: A new framework. OECD Economics Department Working Paper No. 1354.

Egert, B. (2017). The quantification of structural reforms: Extending the framework to emerging market economies. OECD Economics Department Working Paper No 1442.

Furceri, D., Loungani, P. And Ostry, J.D. (2019), The Aggregate and Distributional Effects of Financial Globalization: Evidence from Macro and Sectoral Data. *Journal of Money, Credit and Banking*, 51: 163-198.

Hastie, T., Tibshirani, R., and Friedman, J.H. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer).

International Monetary Fund (2015). Structural reforms and macroeconomic performance: Initial Considerations for the Fund. Staff Report. November 2015.

International Monetary Fund. 2014. World Economic Outlook: Legacies, Clouds, Uncertainties. Ch. 3 “Is it Time for an Infrastructure Push? The Macroeconomic Effects of Public Investment.” Washington, DC, October.

International Monetary Fund. 2019. World Economic Outlook: Global Manufacturing Downturn, Rising Trade Barriers. Ch. 3 "Reigniting Growth in Low-Income and Emerging Market Economies: What Role Can Structural Reforms Play?" Washington, DC, October.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95 (1): 161-182.

Josse, J. and Husson, F. (2013). Handling missing values in exploratory multivariate data analysis methods. Journal de la SFdS. 153 (2), pp. 79-99.

Josse, J. and Husson, F. (2016). missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software*, 70 (1), pp 1-31

Prati, Alessandro & Massimiliano Gaetano Onorato & Chris Papageorgiou, 2013. "Which Reforms Work and under What Institutional Environment? Evidence from a New Data Set on Structural Reforms," *The Review of Economics and Statistics*, MIT Press, vol. 95(3), pages 946-968, July.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011. "From Financial Crash to Debt Crisis." *American Economic Review*, 101 (5): 1676-1706.

Rodrik, Dani. 2010. "Diagnostics before Prescription." *Journal of Economic Perspectives*, 24 (3): 33-44.

Warner, Andrew. 2014. "Public Investment as an Engine of Growth." IMF Working Paper 14/148, International Monetary Fund, Washington.

Appendix

A. List of structural indicators

Table A1. List of structural indicators

	Individual indicator Name
Legal system	Bribery incidence (% of firms experiencing at least one bribe payment request); Bribery index (% of gift or informal payment requests during public transactions); Control of Corruption: Percentile Rank; Firms expected to give gifts in meetings with tax officials (% of firms); Informal payments to public officials (% of firms); Percent of firms expected to give gifts to get a construction permit; Percent of firms expected to give gifts to get a water connection; Percent of firms expected to give gifts to get an electrical connection; Percent of firms expected to give gifts to get an import license;

	Percent of firms expected to give gifts to get an operating license; Percent of firms expected to give gifts to public officials "to get things done"; Percent of firms expected to give gifts to secure government contract; Percent of firms identifying corruption as a major constraint; Value of gift expected to secure a government contract (% of contract value); Business costs of crime and violence, 1-7 (best); If the establishment pays for security, average security costs (% of annual sales); If there were losses, average losses due to theft and vandalism (% of annual sales); Losses due to theft, robbery, vandalism, and arson (% sales); Percent of firms identifying crime, theft and disorder as a major constraint; Percent of firms paying for security; Voice and Accountability: Percentile Rank; Government Effectiveness: Percentile Rank; Intellectual property protection, 1-7 (best); Property rights, 1-7 (best); Efficiency of legal framework in challenging regs., 1-7 (best); Efficiency of legal framework in settling disputes, 1-7 (best); Percent of firms identifying the courts system as a major constraint; Rule of Law: Percentile Rank
Product markets	Burden of government regulation, 1-7 (best); Cost of business start-up procedures (% of GNI per capita); Percent of firms identifying business licensing and permits as a major constraint; Procedures to register property (number); Regulatory Quality: Percentile Rank; Time required to get electricity (days); Time required to obtain an operating license (days); Time required to start a business (days); Time spent dealing with the requirements of government regulations (% of senior management time); Effectiveness of anti-monopoly policy, 1-7 (best); Extent of market dominance, 1-7 (best); Intensity of local competition, 1-7 (best); Firms competing against unregistered firms (% of firms); Firms formally registered when operations started (% of firms)
Financial system	Account at a financial institution (% age 15+) [w2]; Borrowed from a financial institution (% age 15+) [w2]; Debit card (% age 15+) [w2]; Depth of credit information index (0=low to 8=high); Domestic credit to private sector (% of GDP);

	Outstanding mortgage (% age 15+) [w2]; Percent of firms identifying access to finance as a major constraint; Percent of firms not needing a loan; Percent of firms using banks to finance investments; Percent of firms using banks to finance working capital; Percent of firms with a bank loan/line of credit; Percent of firms with a checking or savings account; Private credit bureau coverage (% of adults); Proportion of investment financed by banks (%); Proportion of investment financed internally (%); Proportion of loans requiring collateral (%); Proportion of working capital financed by banks (%); Saved at a financial institution (% age 15+) [w2]; Borrowed to start, operate, or expand a farm or business (% age 15+) [w2]; Debit card in own name (% age 15+) [w2]; Geographical Outreach: Key Indicators, Number of ATMs per 100,000 adults, Number; Geographical Outreach: Key Indicators, Number of commercial bank branches per 100,000 adults, Number; Time to resolve insolvency (years)
Trade and openness	Average time to clear exports through customs (days); Burden of customs procedure, WEF (1=extremely inefficient to 7=extremely efficient); Cost to export (US$ per container); Cost to import (US$ per container); Tariff rate, applied, simple mean, all products (%); Tariff rate, applied, weighted mean, all products (%)
R&D	Capacity for innovation, 1-7 (best); Company spending on R&D, 1-7 (best); PCT patents, applications/million pop.; Percent of firms having their own Web site; Quality of scientific research institutions, 1-7 (best); University-industry collaboration in R&D, 1-7 (best)
Taxation system	Corporate income tax rate, statutory top central; Percent of firms identifying tax rates as a major constraint; Labor tax and contributions (% of commercial profits); Percent of firms identifying tax administration as a major constraint;
Labor markets	Percent of firms identifying labor regulations as a major constraint; Paid annual leave for a worker with 1 year of tenure (in working days); Paid annual leave for a worker with 5 years of tenure (in working days); Fixed-term contracts prohibited for permanent tasks;

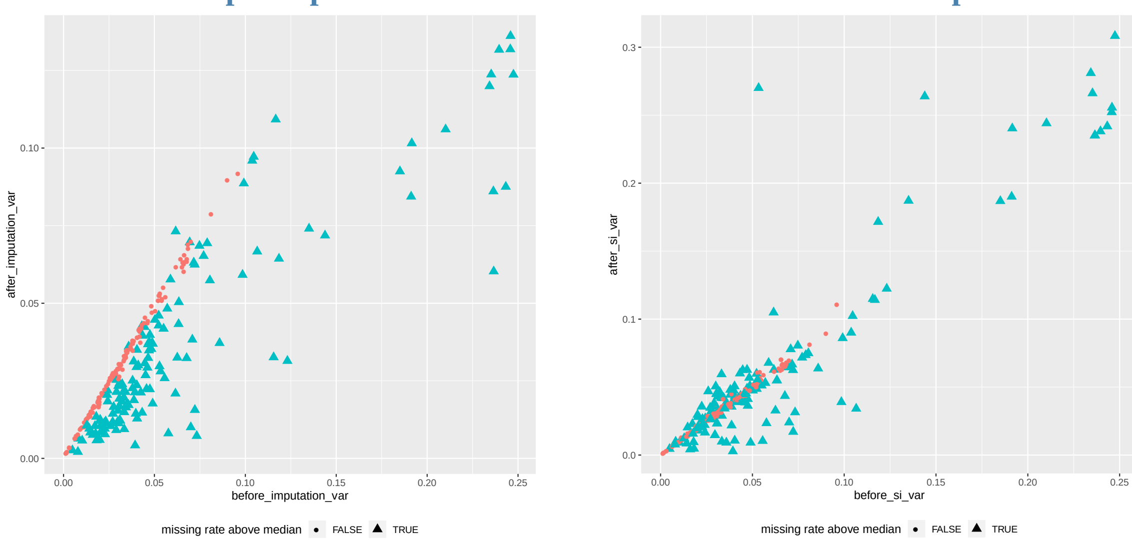
Notice period for redundancy dismissal (for a worker with 1 year of tenure, in salary weeks);
Notice period for redundancy dismissal (for a worker with 10 years of tenure, in salary weeks);
Notice period for redundancy dismissal (for a worker with 5 years of tenure, in salary weeks);
Notice period for redundancy dismissal (weeks of salary);
Paid annual leave for a worker with 10 years of tenure (in working days);
Priority rules for redundancies;
Redundancy costs, weeks of salary;
Severance pay for redundancy dismissal (for a worker with 1 year of tenure, in salary weeks);
Third-party notification if one worker is dismissed;
Maximum working days per week;
Premium for work on weekly rest day (% of hourly pay);
Restrictions on weekly holiday work; Flexibility of wage determination, 1-7 (best);
Minimum wage for a full-time worker (US$/month);
Ratio of minimum wage to value added per worker;
Priority rules for reemployment

B. Description of imputation for missing indicators

We impute the rest of the missing data using a multiple imputation procedure called the iterative PCA procedure (Josse and Husson, 2013).[5] Compared to imputation based on linear interpolation, iterative PCA tries to preserve the variance of each indicator as well as the correlation between indicators during imputation. We can visually assess the variance before and after the imputation in Figure B1. For indicators with a high missing rate, the quality of imputation is lower as reflected in deflated variance. However, compared to linear interpolation, multiple imputation is more robust to missing rates as the variance of the imputed indicator stays aligned to that of the non-imputed indicators, even for indicators with a high missing rate.

Figure B1. Variance of indicators before and after imputation

Multiple imputation **Linear interpolation**



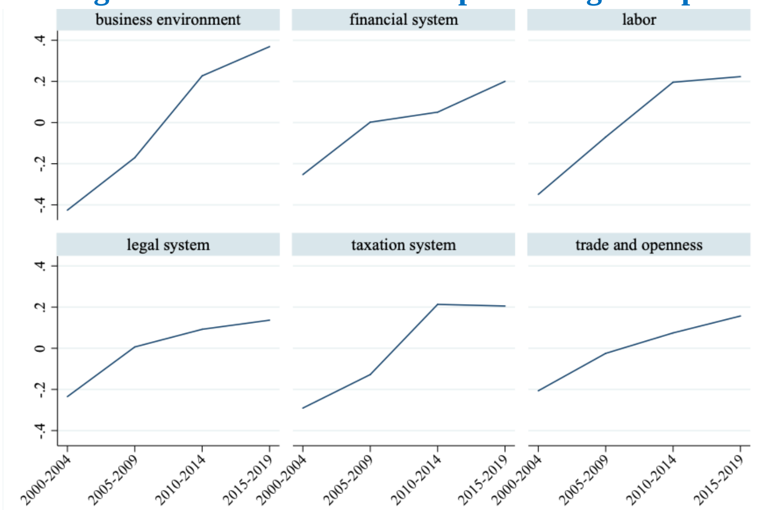
C. Comparison between the PLS structural score and simple-average score

We illustrate the advantage of the PLS model by comparing the PLS score to a simple-average composite. To construct the simple-average composite, within each structural category, for a given country and years, we take the average of indicators in our dataset. As shown in Figure C1, the simple averages would present a different picture regarding the time profile of structural

[5] We perform iterative PCA using the imputePCA package in R. We first rescale all indicators to be between zero and one, and then use the first principal component to impute the missing indicator iteratively. Specifically, at the start of each iteration, this method performs PCA on the dataset with missing values replaced with the mean of the variable plus some Gaussian noise. At the end of each iteration, this method replaces the missing values with the predicted value from the PCA. We standardize the imputed data to be mean zero and standard deviation one.

performance. Different to our metrics, the simple-average composite shows no deterioration in the legal system, while structural performance in the categories of business environment, financial system and taxation system improves more substantially.

Figure C1. Trends of the simple-average composite by structural areas



The horizontal axis indicates the five-year window the structural indicators are collected. The lines plot the average of the simple-average composite across countries. The simple-average composite is the average of structural indicators in a given country and year, standardized to have zero mean and unit variance across all countries and years.

We also repeat the growth analysis in Table 3 using simple averages. As shown in Table C1, the effect of structural reforms on growth becomes statistically insignificant for most structural areas. This is not too surprising and confirms our concern that the simple average assigns larger weights to noisy indicators than the PLS score.

Table C1. Growth impact of structural reforms

VARIABLES	(1) GDP growth	(2) GDP growth	(3) GDP growth	(4) GDP growth
Business environment	0.000919	0.00692	0.000222	8.21e-05
	(0.0211)	(0.0239)	(0.0234)	(0.0204)
Financial markets	0.00810	0.00652	0.0170	0.0105
	(0.0135)	(0.0138)	(0.0126)	(0.0119)
Labor markets	0.00350	0.00394	-0.00818	-0.0125
	(0.0127)	(0.0126)	(0.0136)	(0.0149)
Legal system	0.0319*	0.0283	0.0170	0.0272
	(0.0175)	(0.0189)	(0.0175)	(0.0166)
Tax policy	-0.00225	-0.00183	0.00243	-0.00323
	(0.00653)	(0.00656)	(0.00659)	(0.00551)
Trade & openness	0.0582***	0.0562***	0.0321*	0.0378*
	(0.0189)	(0.0190)	(0.0194)	(0.0206)
Observations	251	251	224	224
R-squared	0.088	0.090	0.295	0.457
Time FE	NO	YES	YES	YES
Country group FE	NO	NO	NO	YES
Controls	NO	NO	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

PUBLICATIONS

Structural Reforms and Economic Growth: A Machine Learning Approach
Working Paper No. **WP/2022/184**