Predicting Wages. A new model to estimate and project wage developments around the globe

Working Paper · March 2016

DOI: 10.13140/RG.2.1.1123.4961

Ekkehard Ernst
International Labour Organization

Christian Viegelahn
International Labour Organization

Predicting Wages*

A new model to estimate and project wage developments around the globe

Ekkehard Ernst†
Christian Viegelahn
Veda Narasimhan
Marta Duda-Nyczak

Winter 2016

Abstract

This paper develops a model to impute missing wage data, and forecast wages over a 6-year time horizon. It uses the relationship between wages and other macroeconomic and labour market variables, to develop multiple model specifications, and then uses pseudo out-of-sample testing to select the best performing models by country. Covering a time period from 1995-2019, this methodology allows for the development of a balanced panel data set for 112 countries. Based on a stylistic framework of wage bargaining, we establish a predictive relationship between a series of covariates and wages. As this model is intended to generate wage data for countries across the globe, we account for the institutional structure of the labour market (collective bargaining vs. individual bargaining) as well as commonalities amongst the economies of different countries such as similarities in the minimum wage law. The novelties in our methodology however is the use of vulnerable employment and non-working poverty, to proxy for the outside option in our wage bargaining framework, and the use of out-of-sample testing, in order to cross-validate the results of our specifications. These additions allow us to develop forecasts on the basis of model averaging techniques using an approach that is specifically tailored to minimize the forecasting error at the country level.

*The work behind this project started as an attempt to provide complementary labour cost information for the ILO's Global Employment Trends model. Our special thanks go to colleagues from the ILO's INWORK unit, in particular Patrick Belser, Kristen Sobeck and Rosalia Vazquez-Alvarez who have provided invaluable input and regular data updates that allowed us to proceed with the estimates in this paper. Several colleagues in other organisations have also provided comments and input at various stages, in particular Prakash Loungani (IMF), Bart van Ark, Gad Levanon, Bert Colijn (all The Conference Board) and participants at the *Workshop on International Labor and Productivity Comparisons*, Washington, DC (October 2015).

†All contributing authors are – or have been in the past – members of the Research Department of the International Labour Organization. Ekkehard Ernst: ernste@ilo.org; Christian Viegelahn: viegelahn@ilo.org; Veda Narasimhan: narasimhan@iloguest.org; Marta Duda-Nyczak: MDudaNyczak@uneca.org.

1 Introduction

By now, wages are the most frequent form of labour income in many countries of the world. Worldwide, the share of salaried workers has been on the rise, as opposed to the share of workers in vulnerable employment. In 1991, less than 41 per cent of all workers in the world were in wage and salaried employment, a number which is estimated to have increased to almost 53 per cent in 2015. The increasing importance of salaried employment is a phenomenon that is prevalent in a majority of countries across all regions. With more than half of the global worker population in wage and salaried employment, wages are becoming a more and more representative indicator of the economic well-being of workers worldwide.

In most cases the wage level determines the living standard of the worker. Therefore, the average monthly wage in an economy constitutes a non-exclusive but important indicator of decent work and material progress of the working population. On the other hand, paying wages contributes to an employer's production costs and has a considerable impact on the economic viability of companies, especially in times of fierce international competition. Having access to timely, reliable and comparable information on wages is crucial to inform the wage-setting process between social partners who should both ensure that decent work conditions are implemented.

Wages also play a crucial role for economic governance as they have important implications for economic growth. Wage policies, especially in countries with current-account surpluses, can contribute to a strengthening of domestic demand, thereby reducing imbalances in the global economy. And indeed, recent empirical evidence suggests that wage policies can have a considerable impact on output through the demand channel, generating large multipliers and bringing an economy back on a sustainable growth path (Charpe and Kühn, 2015). On the other hand, some economists argue that under certain circumstances a strong increase in wages can reinforce recessions, bringing an economy off the growth track. Given the importance of wages in the macroeconomic context, wage data and forecasts crucially inform government choices regarding economic and labour market policies and can herewith contribute to better economic outcomes.

For this purpose a wage database with real-time and balanced information is indispensable. In particular, interest in wage forecasts is strong as they provide an indication on world and regional trends of wages. Without a consideration of such trends, future developments of labour market indicators such as working poverty or employment are difficult to assess. For these reasons, the ILO has put significant emphasis to develop a sound and well-founded methodology that can be used for the imputation of missing wage data and the production of wage forecasts (see ILO, 2014, appendix I). Despite the importance of understanding wage trends globally, there have been only relatively few attempts at producing an econometric model to generate predictions of wages. This paper seeks to introduce a new methodology that will allow for imputing missing wage data, in addition to generating predictions of real wage trends for 112 countries.

The Global Wage Database (GWD) of the ILO's Inclusive Labour Markets, Labour Relations and Working Conditions Branch (INWORK) contains most of the wage data series that are available via ILOStat and were selected based on the criterion of comparability.[1] We convert the nominal wage series in local currency units into real wage series in 2005 international dollars at PPP ($) by using data on consumer price inflation and the PPP conversion factor from the IMF. We end up with wage data for 112 countries of which 75 are from the developing world. This corresponds to 63 per cent of countries and 53 per cent of developing countries that are included in the annual ILO's World Employment and Social Outlook, 2015 (ILO, 2015).

However, especially for developing countries the wage series contain significant gaps and the available time series are short. Especially, differences in coverage across regions are huge. While coverage ratios for Central and Southeast Europe mostly reach levels of above 70 per cent, coverage ratios for Sub-Saharan African countries usually do not exceed 20 per cent in our input database. Coverage ratios for Latin America and the Caribbean and North Africa are always below 50 per cent, while those for the Middle East, East Asia, Southeast Asia and the Pacific, and South Asia reach levels of more than 50 per cent, at least in some years.

This paper develops a methodology that can be used to impute missing wage data at the country level and develop real wage forecasts over a short- and medium-term horizon (up to 5 years). A particular challenge to fill the data gaps relates to detecting the patterns at which data are missing. Here, we distinguish between wave and unit non-response. While wave non-response describes the situation in which some data are available for a country, though not for all years, no data at all are available for a country when we speak of unit non-response. The two missing data patterns require different treatments. In this paper, we develop a model-based methodology that is suitable to impute missing data and helps generate forecasts for those countries, for which two consecutive wage observations are available. Hence, we provide a methodology for wave non-response missing patterns.

We introduce a methodology that is able to produce wage estimates and forecasts not only, but in particular, in a developing country context for countries and years in which actual data are not available. The estimations perform well compared to several benchmark models. The imputation methodology relies on a theoretical framework linking outside labour market conditions to wage outcomes. Such a relationship can be derived from a wage curve approach, assuming that the capacity of workers to earn higher wages is larger under more favourable economic and social conditions (see Blanchflower and Oswald, 1994). In such a framework, wages can be derived as a function of expected productivity and expected reservation utility, where reservation utility is defined as the utility in monetary terms that workers obtain when negotiations with the employer break down and do not result in an employment contract. In most developing countries and in the absence of well developed social protection systems, the reservation utility is determined by the level of poverty among those that are not in wage and salaries employment. The lower the fallback option for workers is - i.e. the more widespread non-working poverty is in a country - the less workers will be able to share in any productivity gains. In contrast, when non-working poverty is absent or at least limited, then workers have much more possibilities to achieve larger wage gains

[1] Available at: `https://www.ilo.org/ilostat/`

either through wage bargaining or by switching jobs to higher paying alternatives.

In order to ensure a sufficiently high quality of the estimates, wage estimates and imputations need to rely on the most comprehensive set of relevant and readily available information, combined with a sound theoretical foundation. In this paper, we use labour productivity, working and non-working poverty rates among those that are not in salaried and wage employment, the share of vulnerable employment and investment as a share of GDP as explanatory variables for wages. Based on these input data, we produce wage estimates for a balanced panel of 112 countries between 1995 and 2014 and produce wage projections up to 2019.

The paper is organised as follows: The next section provides an overview of the literature on the relationship between wages and other macroeconomic variables, and presents previous attempts at forecasting wage data. Section 3 then presents the underlying theory and provides the introduction into the bargaining framework that underlies our estimation methodology. Section 4 derives how the theory can be empirically applied in order to produce wage estimates, Furthermore it discusses the data needed for this task and the estimation methodology that is applied. Section 5 looks at the regression results and presents our novel testing approach for selecting our best-performing models. Furthermore, based on the predictions of our preferred model, we derive imputed values for missing wage data and evaluate the wage estimates using cross-validation techniques to asses the reliability of our imputation and projection methodology. A final section concludes.

2 Literature Review

Few models have attempted to impute and project missing wage data, varying in terms of the country coverage, model specifications and level of analysis conducted. Despite an abundance of papers dedicated to analysing the interaction of wages and employment, there still remains some debate on the mechanism through which wages adjust, and the variety of methodologies chosen in these wage projection models reflects this diversity.

The private audit company PriceWaterhouseCoopers (PwC) produced the *Global Wage Projections to 2030* (Hawksworth and Lambe, 2013), constructing projections on the basis of the relationship between labour productivity, exchange rates, and real wages. The coverage of countries in this study however is more limited, and the focus of this report is the wage gap between developed and emerging economies. The projection period, however, is much longer, running until 2030. As in our paper, the model relies on the ILO Global Wage Database. Forecasts are adjusted downwards for advanced economies to account for the recently weaker relationship between real wage growth and labour productivity growth.

The Conference Board, a private business membership and research association in the United States, takes another approach, looking at labour cost projections for OECD countries (Levanon *et al.*, 2016). This method produces labor cost projections on a nominal basis, accounting for the impact of macroeconomic variables such as inflation, unemployment and employment protection.

This model covers a panel of 19 advanced economies, taking advantage of better data availability in these countries.

The ILO's INWORK team has developed an econometric model to produce regional averages based on a weighting scheme of available country-level information. This model is based on a probit estimation that determines factors that explain the probability of a country to not report any wage data, which is used to weigh the reported data points when aggregating wage figures to regional wage averages. Hence, whereas country-level information in the Global Wage database reports only (officially) available data, regional averages allow to detect longer term trends for each of the six regions and the world as a whole.

Although there is limited research into methodologies to account for missing wage data, and to forecast future wages, much research has been done on the relationship between wages and other macroeconomic variables. Given this debate, the model presented in this paper does not seek to add to this debate but rather develop a way to capture the basic effects of this mechanism. The common theme between both wage forecasting models presented above is that they use the relationship of wages to other macroeconomic variables in order to generate projections of the future trends in wages. The rest of this literature review will deal with looking at the links presented in the literature between wages and other macroeconomic variables.

Standard macroeconomic theory relates wage growth to unemployment using different variations of the Phillips curve argument. The Phillips curve relates real wage *inflation* to unemployment, stressing that wage claims can be more easily achieved in periods of low unemployment rates. Hence, wage growth is negatively related to the level of the unemployment rate. On the basis of micro-econometric evidence, Blanchflower and Oswald (1994) refute this relationship and claim that it is instead the *level* of wages that is inversely related to the unemployment rate, giving rise to the "wage curve". Similar to the wage Phillips curve, however, it is changes in the outside option for workers that allow them to achieve higher wages in periods and areas where unemployment is low. The wage curve makes up for a more persistent impact of (supply) shocks on wages as workers will aim at recovering the (expected) level of wages even after the shock has disappeared. Even though both claims look irreconcilable, Whelan (2000) argues that the the wage Phillips curve is compatible with any micro-economic relationship between the level of wages and unemployment. However, existing evidence points to a more persistent effect of shocks on wages, which is also reflected in our empirical strategy discussed below.

Bargaining theory has been applied to a wide range of topics, and has been used extensively in wage determination theory (see, e.g., Binmore *et al.*, 1986; McDonald and Solow, 1981). The basic framework involves a (representative) worker and a (representative) employer, who determine the optimal wage level through a bargaining process. Wage outcomes differ depending on the specifics of the bargaining process. For instance, wage bargaining power is weaker when several (competing) unions are present rather than one unified one. The simplest bargaining framework has a single monopolistic union that sets the average wage, and a single firm that then selects the employment level (the 'right-to-manage' model, see McDonald and Solow, 1981). An alternative and more commonly observed model deals with a union that gains utility from members as well as wages, and a firm that aims to maximize profits. This changes the interaction between the two actors, thereby changing the wage determination process. The inclusion of variables to account for the wage bargaining process is justified by the idea that there is a differing impact on wage growth (or levels) when the bargaining systems differ. Calmfors and Driffill (1988), for instance, argue that there is a hump-shaped relationship between the degree of centralization and the real wage.

In addition to considering wage bargaining at the macroeconomic level, many papers also study the impact of wage bargaining at the sectoral level. The argument is that as important as productivity and unemployment levels are to determining wage levels, industry growth and employment also play an important role in determining the aggregate wage level and growth rate (Dreger and Reimers, 2010). These models however tend to deal with analysing the effect of idiosyncratic elements that operate through the sectors.

3 Wage determination

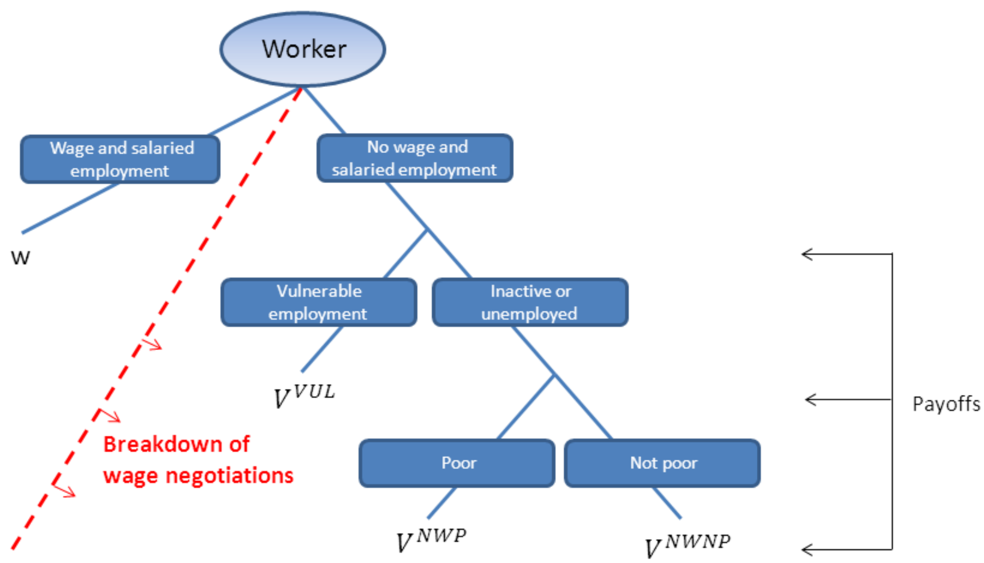
3.1 Overview

The literature review in the previous section suggests that (individual) wages react - to a certain extent - to economic circumstances of the wider economy. Labour incomes - whether paid in the form of regular salaries or gained as sporadic payments for temporary and self-employment - vary depending on the dynamics of aggregate demand, the capital intensity of an economy and the number of other people on the labour market looking for jobs or better employment possibilities (i.e. unemployed and potential job-hoppers). In particular in emerging and developing countries, this latter category represents the vast majority of labour market participants as open unemployment is very small and mostly limited to certain fortunate job seekers who can afford to remain unemployed thanks to private resources.

The theoretical model we develop in this section aims at reflecting this. It starts from a conceptual framework with bargaining over wages. Such bargaining can take place both at the collective level (national, sectoral, firm-level union bargaining) and - most often, given the limitd representativeness of organised labour in many parts of the world - at the individual level (with obvious limitations to workers' bargaining power). Besides the economic situation of the firm (i.e. its profitability), the outcome of the bargaining process will depend to a large extent on the outside option (i.e. the alternative employment opportunity) of the worker.

Absent proper social protection systems, a worker's outside option can take three states (see figure 1): (i) an employment in a vulnerable employment situation either as own-account worker or as a contributing family worker, V^{VUL} (ii) inactive or unemployed non-working poor, $V^{NWP}$and (iii) inactive or unemployed non-working-not-poor, V^{NWNP}. For the last two situations, we do not make a distinction between inactivity and unemployment, using instead the ILO's new concept of the potential labour force that includes also those who are currently not actively seeking a job but willing to take one up if one would become available.[2]

Figure 1: Workers' inside and outside options



3.2 A theoretical model

Starting from these conceptual considerations, we set up a formal model to illustrate the mechanism of the wage bargaining process. This helps us in setting up the empirical model specifications we want to estimate in the next section. In the model, we assume that we can determine (theoretically) a representative wage for the whole economy through bargaining between employers (firms) and workers. The framework is general enough to include also cases in which employers set wages unilaterally. We start by looking at the firm's maximization problem:

$$E\left(\pi\left(w,L\right)\right) = E\left(R\left(L\right)\right) - w \cdot L$$

Here we assume that the firm is a profit-maximizer, with a fixed capital input. Given that capital and investment decisions are more long-term, we can assume them as fixed at the point in time

[2]The concept of the potential labour force was discussed and adopted by the 19th International Conference of Labour Statisticians in 2013: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf.

when the bargaining starts. In this situation, the employers' budget constraint problem is limited to the labour decisions. The term $E(R(L))$ gives the expected revenue of the firm, as a function of labour (effectively labour productivity).

In this bargaining model we have a single worker representative (which could, for example, be a trade union) that bargains with the employer representatives in the economy. We assume that the worker representatives have Stone-Geary type preferences, which can be represented as a utility function that accounts for within-union bargaining (Geary, 1950; Stone, 1954). This function represents the product of this bargaining between the union that has a preference for members, and the members who derive their utility from wages. This gives us the following expected utility function:

$$E(u(w,L)) = (w - E(v))^{\theta} L^{\gamma}$$

In this equation, w is the wage level, v is the outside option and θ and γ are the relevant elasticities. We can view $E(v)$ as the expected valuation of the outside option available to the individual in the event of bargaining failure. Given that the outside option has a certain amount of risk associated with it, this value is not realized until after the bargaining stage, hence it enters the function as an expected value. The Nash equilibrium can then be found by maximizing the following expression for the aggregate economy:

$$\Omega = [E(\pi(w,L))]^{1-\beta} [E(u(w,L))]^{\beta}$$

The β in this equation represents the bargaining power of the union versus the firm, where $\beta \in [0,1]$, where 0 is complete bargaining power with the firm (corresponding to a situation in which the firm unilaterally sets the wage), and 1 represents full bargaining power with the union (corresponding to a situation in which worker representatives unilaterally set the wage).

Solving the maximization problem with respect to wages and labour gives us the following two first order conditions:

$$\Omega_w = \Omega \left[\frac{\beta\theta}{w - E(v)} - \frac{(1-\beta)L}{E(R) - wL} \right] = 0$$

$$\Omega_L = \Omega \left[\frac{(1-\beta)\left(\frac{\partial E(R)}{\partial L} - w\right)}{E(R) - wL} + \frac{\beta\gamma}{L} \right] = 0$$

Using these two equations we can derive the contract condition which gives us a set of points for which the union's indifference curve is tangent to the firm's isoprofit curve:

$$\frac{\gamma}{\theta}(w - E(v)) = w \cdot E(v) - E(R_L)$$

By manipulation of the first-order condition for wages, we get the rent division (Nash-bargaining) condition:

$$w = \frac{\beta\gamma}{1-\beta+\beta\gamma} E(A) + \left(1 - \frac{\beta\gamma}{1-\beta+\beta\gamma}\right) E(R_L)$$

In this condition, $E(A) = \frac{E(R)}{L}$ is the expected average labour productivity, and where $E(R_L)$ is the expected marginal productivity of labour. Substituting in the contract condition, and using the first-order condition for labour, we end up with the following result:

$$w = \frac{\beta\theta}{1-\beta+\beta\theta} E(A) + \frac{1-\beta}{1-\beta+\beta\theta} E(v)$$

This result suggests that wages are the weighted average of expected reservation utility of the unions, and the expected average productivity in the economy.

If represented at the sectoral or occupational level, the wage bargaining model can also account for differences in wage setting across sectors or occupations. If there is a huge wage differential between sectors and occupations, the distribution of employment across sectors and occupations will naturally affect the average wage in the economy. For example, a rapidly growing high-skilled sector in which higher wages are earned will push the average wage level in the economy upwards, whereas a growing low-skilled sector will move average wages in the opposite direction. Understanding the distribution of employment in the economy across different sectors and occupations is crucial to determining the mean wage level, given the different weights that different sectors contribute to the general economy.

4 Methodology

The forecasting model that we develop in this paper is based on the idea that we can make use of empirical relationships between wages and a series of macroeconomic and labour market variables for which we have predictions available for a 6-year forecasting horizon. We regress our dependent wage variable on combinations of independent variables, using different panel estimation techniques, country groupings and specifications. The results of these regressions provide us with coefficients that measure relationships between wages and the independent variables based on past data. We then apply these coefficient estimates to forecasts of our independent variables to give us a prediction of our dependent wage variable. Given the structure of the model, we are restricted by our choice of independent variables, as we can only select those for which we can get forecasts for the same time horizon.

4.1 Data

The main source of wage data in this model is the ILO's Global Wage Database (ILO, 2014). This database consists of an unbalanced dataset of monthly wage data in local currency. Many countries provide multiple data series for wages, for example one series for urban jobs and one for rural jobs. As much as possible we have attempted to keep consistency across all series, by choosing series that cover a wide range of sectors and regions within a country, and therefore are representative of the labour market as a whole. In addition, we have tried to select series that have a longer time series available, and at least two consecutive actual data points, in order to be able to calculate growth rates.

The panel runs from 1995 to 2014, although the bulk of the data covers the period 1999-2014. As we are looking at annual real wage growth, we first convert our data series to an annual series. The original data are reported in nominal terms, so we deflate the data series with consumer price inflation data taken from the IMF's World Economic Outlook (WEO), and then convert it using the relevant PPP conversion factor so that all our wage series are in 2005 international $ (PPP), also taken from the WEO. This gives us an annual real wage series for all the countries that have sufficient data available. Since we are interested in developing wage growth numbers, we are limited in the data we take from the GWD, in that the series has to have at least two consecutive data points in order to be included in our series.

In order to construct labour productivity measures, we use data from the World Bank's World Development Indicator (WDI) Database on GDP in US dollars ($), and employment data as produced by the ILO's Global Employment Trends (GET) model as discussed below. By dividing the GDP data by the employment numbers available, we get a labour productivity series for 1991-2019. In addition, we use IMF WEO data on investment as a share of GDP, which is forecasted till 2019, thereby providing us with a full set of input data until 2019. As the IMF updates this data series twice a year, these wage numbers can be regularly updated using the newest WEO numbers, as well as using the annual WDI update. Due to the high volume of data involved in this model, this update is usually done only twice a year.

Employment data used in the model comes from the ILO's Global Employment Trends (GET) model. The GET provides a fully balanced database of employment, unemployment, sectoral employment and occupational employment data for 178 countries from 1991-2019, based on data being sourced from governments through population censuses, household surveys and labour force surveys. Missing data points are imputed by the model, and employment is forecasted through 2019. A crisis dummy and a recovery dummy is also collected from this database, and gives us a dummy variable that accounts for the years when individual countries were in recovery or in crisis in relation to the 2007-2008 global economic crisis. The data included are as follows: unemployment rates, total employment, vulnerable employment, wage and salaried employment, employment in broad sectors (agriculture, services and industry), employment in disaggregated sectors[3], employment by occupational skills (low skill, medium skill, high skill), employment by

[3]The fourteen sectors covered in the ILO GET model are: Agriculture & forestry & hunting & fishing, Mining

type of occupational tasks (routine, non-routine manual, non-routine cognitive).[4] Although we collect data on all these categories, we only include some of the variables in our specifications in order to prevent collinearity.

Other explanatory variables include unemployment duration taken from the Key Indicators of the Labour Market (KILM) database, which provides information on the average number of months of unemployment for countries.[5] In addition, we use ILO data on working poverty rates, for multiple categories to calculate, in combination with information on national poverty rates from PovcalNet, non-working poverty rates which correspond to the share of workers that are not in wage and salaried employment and poor. We use the measure of less than \$2 a day as our benchmark for poverty. Moreover, we include information from the ILO Working Conditions Laws database on the minimum wage fixing mechanism and minimum wage coverage in order to create country groupings on the basis of labour market mechanisms.[6] Finally, we use data on the size of the population from the UN World Population Prospects as inputs into our model, to control for demographic effects on the wage level in an economy. For an overview of the different variables and their sources, see table 1.

Table 1: Independent Variables and their sources

Independent Variable	Source
Productivity	World Bank; ILO GET
Investment to GDP ratio	IMF WEO
Youth-to-total Population Ratio	UNWPP
Working-Age to Total Population	UNWPP
Crisis and Recovery Dummies	ILO GET
Vulnerable Employment (Total and Share in Non-Salaried)	ILO
Non-working and Working Poverty	ILO
Broad Sector Employment	ILO GET Extension
Disaggregated Sector Employment	ILO GET Extension
Occupation by Status Employment	ILO GET Extension
Occupation by Skill Employment	ILO GET Extension

& quarrying, Manufacturing, Utilities (gas and electricity), Construction, Wholesale & retail trade & transport repair, Accommodation & restaurants, Transport & storage & communication, Financial activities, Education, Health & social work, Public administration & defence & compulsory social security, Real estate & business and administrative activities, Other.

[4]The classification of occupational employment by tasks follows Autor *et al.* (2003).

[5]The ILO KILM database is available at `http://www.ilo.org/global/statistics-and-databases/research-and-databases/kilm/lang--en/index.htm`.

[6]ILO Working Conditions Laws Database. ILO, Geneva. Available at: `http://www.ilo.org/dyn/travail`

4.2 Estimation Techniques

The model presented relies on three different panel estimation techniques: fixed effects (FE), mixed effects (ME) and generalized least squares (GLS). This mix of estimation methodologies, allows for a mix of representative models by country, so differences in the wage-setting mechanism can be reflected by different methodologies.

We use the FE estimator in order to control for unobserved heterogeneity across countries that is fixed through time. Country fixed effects control for time-invariant country-specific factors influencing wages, which the estimator allows to be correlated with the other independent variables. The GLS estimator allows us to account for idiosyncratic errors that are correlated through time (Duncan and Leigh, 1980). In addition, this estimator allows for heteroskedasticity in the data, producing unbiased coefficients.

We also use ME estimators that include components of both random and fixed effects. While fixed effects are estimated directly, random effects can be generated throughbest linear unbiased predictions. ME models allow for heterogeneity of coefficients across countries, which can be important in the context of our model. These models are particularly useful in the context of panel datasets, especially when a large number of observations are missing. ME models allow us to account for unobserved heterogeneity of coefficients across different countries.

We also control for business cycle movements by introducing a dummy variable for crisis and recovery periods. As described earlier, this dummy indicates the crisis (recessionary) and recovery periods for every country. This helps us to control for wage effects determined by the business cycle rather than firm expectations, and shocks to the whole economy, that may not necessarily have been accounted for in the expectations set at the beginning of the period (for example, a sudden downturn in the economy). The inclusion of these variables is useful in order to ensure that the model does not include business cycle effects in the coefficient estimates, which would give us biased predictions and imputations. In addition, in order to control for interactions between our independent variables, we also include interaction terms between variables included in our specification, year interaction terms, and crisis interaction terms.

4.3 Country groupings

In order to control for effects specific to certain groups of countries, we run the regressions on a series of 6 different groupings. First, we run regressions on the whole sample. This provides us with the highest number of observations, but averages out any country-specific effects. This grouping gives us a global specification which can be applied to all countries. Second, we group countries geographically, in order to account for regional specificities of the wage setting mechanism. This grouping is not strictly regional however, as Advanced Economies are grouped in a single category, which allows us to capture specificities of this economic group. Next, we introduce two types of country groupings on the basis of the minimum wage legislation, in order to account for the impact of labour market mechanisms on wage determination. First, we construct country groupings based on whether minimum wages are set nationwide, by sector, or do not have a minimum wage. Second, we group countries according to whether minimum wages are set by the government or a specialized body, through collective bargaining, or whether no information are available, or no minimum wage is in force. The reason for including countries that have no data into a single category, is to account for underlying similarities in these countries: for example, many poor Sub-Saharan African countries do not have the institutions in place to track these data, but these countries have a lot of economic characteristics in common, which make them good candidates for a single group.

Our final two groupings deal with unemployment duration - we gather these data from the KILM database. The first of these unemployment categories simply group countries together by the average duration of unemployment in months across all years: category 1 consists of countries with an average number of months between 0 and 9, category 2 consists of countries with an average number of months greater than 9, and the final category deals with countries that report no data. In order to calculate the average unemployment duration, we first take data on the share of unemployed by year in the categories "less than 1 month", "1 month to 3 months", 3 months to 6 months", "6 months to 12 months" and "More than 12 months". We then weight each category by its midpoint (for example, the share of those in the "less than 1 month category" gets multiplied by 0.5) and sum them up to get a single average for the whole country for that year. For the final category "more than 12 months", we use 18 months as the midpoint. We then take an average by country to get a single value for every country. This is the value used to assign countries to the categories describer earlier.

The final grouping deals with the pre-crisis to post-crisis ratio of unemployment duration. The pre-crisis period is between 2000-2007, while the post-crisis period is 2008-2012. Using the weighted average duration for each year described above, and average across the pre- and post-crisis period highlighted. We then divide the post-crisis average by the pre-crisis average, and sort countries according to those who had a ratio greater than 1.2, those that were less than 0.9, and those that were in between. This allows us to capture the impact of the crisis on the labour market, by sorting countries according to their ability to recover from the effect of the crisis. This brings the total number of groups to 6, with each of the last 4 categorized into 3 sections.

All country groupings aim to account for similarities and wage setting institutions across countries, allowing estimated coefficients to vary across countries that belong to different groupings. By accounting for heterogeneities of countries in employment legislation, labour market turnovers, and geographical proximity, we can capture different effects of the complex relationship between wages and the economy as a whole, alongside capturing country-specific characteristics, that a general specification would average out.

4.4 Dependent variable choice

Given our theoretical model, we start constructing different possible models to determine wage growth in the economy. Our base model uses the growth in real wages as the dependent variable. This is our main variable of interest, and in the final stages of our model, all wage data is provided in this form (imputations and predictions). This version of our dependent variable allows us to use the direct relationship between wages and productivity, and use it to forecast wage growth directly. In order to control for a series of different interactions between the variables however, we extended the model to include two alternative dependent variables from our original one: the log of the level of wages, and the difference between the growth in wages and growth in productivity. This last variable is based on the assumption we made earlier in the paper that productivity and wages move closely. Each dependent variable has its own set of specfications, which do no necessarily repeat. We also go through these models, checking the results of the coefficients, and dropping model specifications that are not suitable for our purposes.

The reasoning behind selecting alternative dependent variables, is to try and capture different sets of effects, and differences in the relationship between wages and other macroeconomic variables. By using alternative dependent variables, and pairing them with modified independent variables (for example, with the log of real wages we use the log of productivity), we are able to analyze different wage-determining relationships. Additionally, we capture both level effects and growth effects of wages through two of our dependent variables, while our dependent variable based on the difference between wages and productivity allows us to track the evolution of this relationship, which allows for some flexibility in the relationship between these two variables.

4.5 Model Specification

We use the theoretical model presented earlier in this paper to give us a starting point for our empirical model. The equations presented earlier gives us two possible models for our empirical wage equation; the first describes the wage as a weighted average comprising of the expected reservation utility of the union and expected labour productivity. As the purpose of developing this model is to develop the best possible forecasts, we develop a whole series of possible regression models to cover different interactions. The main limitation to our choice of variables is the need to have a large set of past data available, as well as a set of forecasts. We first considered the traditional bargaining model at the national level.

We set up our wage equation with time-specific variables, given that we are looking at a panel dataset. It makes sense here that the expected variables are lagged by one period, as individuals can only use information up to the previous year as a basis for their expectations for the next year. This gives us an estimation equation of the following structure:

$$w_t = \frac{\beta\theta}{1-\beta+\beta\theta} E_{t-1}\left(A_{t-1}\right) + \frac{1-\beta}{1-\beta+\beta\theta} E_{t-1}\left(v_{t-1}\right) \tag{1}$$

It is obvious that we do not have a measure for expected labour productivity, rather we proxy for this variable using the ones we have available, more specifically variables that individuals would use when creating their expectations. It seems reasonable to assume that individuals base their expectations of the future on past information, but how many periods of information to use is an empirical question, and we experiment with different lags. The caveat in this case is that every additional lag we add reduces the wage database by one year of wage information and hence reduces the number of observations we can use. In addition to past information on productivity, we also include the lagged investment share of GDP as an additional determinant of expected labour productivity. Investment has a positive effect on productivity, and an increase in investment as a share of GDP in the previous year is likely to send a signal to firms of higher productivity in the following year.

In order to construct a measure of the expected reservation wage of the worker representative, we use data on employment by status. In the bargaining model, someone who refuses a specific wage or salaried employment position, can end up in one of three (technically four) different situations. First, they could end up in vulnerable employment as working poor, second they could end up in vulnerable employment as non-working poor, or third (and fourth) they could end up unemployed or inactive, but for the purpose of this paper, these two are assumed to be one category. It can be assumed it is preferable to be in vulnerable employment rather than unemployed, as in the case of vulnerable employment at least there is still a source of income. In addition, vulnerable employment also accounts for own-account workers and contributing family members, a category that can include highly successful entrepreneurs as well. In all these cases a breakdown of employment by status is required, and the GET model data provides us with the perfect source of input data. To model the expected utility from the reservation wage, we need to find the probability that an individual is left unemployed or in vulnerable employment. We also need to distinguish between the poor and non-poor of those who are not in wage and salaried employment. For this we use the vulnerable employment as a share of total employment, and share of non-working poor as a share of those not in wage and salaried employment, to proxy the probability of ending up in either of these situations.

Although using the full balanced panel dataset of vulnerable employment gives us more predictive power, it increases the statistical error of our models as a large portion of this data is estimated. In order to prevent this, we restrict our data sample to real values of vulnerable employment instead to drive down this error. Although this reduces some of the R^2 of our model, it allows us to ensure that the statistical error on our model declines.

In addition, we can use the above proxy variables for labour productivity (which we have assumed to be consistent across all sectors), but we need to find a measure for the probability that an individual ends up in a different sector, which we proxy by using the share of employment of that sector or occupation of total employment and estimating a slightly modified version of (1):

$$w_t = \frac{\beta\theta}{1-\beta+\beta\theta} E_{t-1}\left(A_{t-1}\right) + \frac{1-\beta}{1-\beta+\beta\theta} E_{i,t-1}\left(v_{\neg i,t-1}\right)$$

where i : index of economic sector and $E_{i,t-1}(v_{\neg i,t-1})$: the expected probability for a worker in sector i to find an alternative employment in a different sector ($\neg i$).

However, there is an endogeneity problem associated with determining the average wage level through sectoral wages (Lee and Pesaran, 1993): The average wage level influences the sectoral wage level, whereas the sectoral wage level determines the average wage level. In order to get around this issue, we use the lagged employment share by sector. We include models that account for growth and levels of employment in a range of sectors. We also use different levels of aggregation, ranging from 3 main sectors to 14 disaggregated sectors.

A similar issue occurs when using occupational data, both by skill level and status level. As before, these enter our specification with a lag. This allows us to observe the effect of these different categories on the average wage level in the economy. The evolution of these categories can provide interesting insights into the evolution of wages in turn; for example, countries that have rapidly growing high skill sectors are likely to witness upward pressure on the wage level, from the higher wages paid in this sector. These specifications allow us to consider countries with differing patterns of development, and observe the likely evolution of their average wages over the next few years.

In summary, we first run a series of specifications that make use of different indicators of the outside option (vulnerable employment, non-working poverty, working poverty), alongside labour productivity and the investment share. We then proceed to test sectoral level effects, by running labour productivity and the investment share, with employment in different sectors, both broad and disaggregated (as presented in the GET model). Finally, we run labour productivity and the investment share alongside occupational data (low skill, medium skill and high skill), and occupational status (manual routine, manual non-routine, non-routine cognitive). These different combinations of variables, capture different aspects of the labour market, which are reflected in the average wage. For example, countries with high levels of low skilled workers, will have a lower average wage level, as low skilled workers are paid less. These specifications are also run with control variables for population characteristics, such as working age-to-total population ratio, and youth-to-total population ratio. Additionally, many of the specifications make use of the crisis or recovery dummy discussed earlier, to account for business cycle fluctuations in the wage level.

Overall, we run approximately 900 different types of regressions in total for our model, not all of which go into our final model selection. This includes three different specifications of the dependent variable, where we us real wage growth, the difference between real wage and productivity growth and the logarithmic level of real wages. We combine these with various combinations of independent variables in addition to running the same specification on different groupings and making use of different estimation methodologies (GLS, FE and ME). This gives us a large pool of regression results to analyse. The following tables 2-4 give an overview over the different specifications retained for our regression analysis.

Table 2: Specification overview: Real wage growth

Dependent variable: Δw_t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ΔLP_{t-1}	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
ΔLP_{t-2}	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$\Delta InvSh_{t-1}$	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$\Delta YouthRatio_t$	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$\Delta WPopSh_t$	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$VulnEmpSh_{t-1}^{non-salaried}$		■	■			■	■	■	■	■	■	■								
$NonWorkPoorSh_{t-1}^{\$2}$		■	■						■	■		■								
$\Delta VulnEmpSh_{t-1}^{non-salaried}$				■	■															
$\Delta NonWorkPoorSh_{t-1}^{\$2}$				■	■															
$WorkPoorSh_{t-1}^{\$2}$						■					■									
$I_t^{WorkPov \times VulnEmpSh^{non-salaried}}$						■														
$WorkPoorRatio_{t-1}^{\$2}$							■	■												
$I_t^{WorkPovSh \times VulnEmp^{non-salaried}}$							■	■												
$DepRatio_{t-1}$								■												
$I^{Crisis} \times \sum_i IndepVar_i$										■	■									
$Year \times \sum_i IndepVar_i$												■								
$\Delta EmpSectorSh_{t-1}^{Broad}$													■							
$EmpSectorSh_{t-1}^{Broad}$														■						
$\Delta EmpSectorSh_{t-1}^{Detail_{14}}$															■					
$\Delta EmpSectorSh_{t-1}^{Detail_4}$																■				
$EmpSh_{t-1}^{Occupation-by-task}$																	■			
$\Delta EmpSh_{t-1}^{Occupation-by-task}$																		■		
$EmpSh_{t-1}^{Occupation-by-skill}$																			■	
$\Delta EmpSh_{t-1}^{Occupation-by-skill}$																				■
I^{Crisis}	■		■		■	■	■	■	■	■	■		■	■	■	■	■	■	■	■
$I^{Recovery}$									■											

Note: Δw_t - Real wage growth, ΔLP_{t-1} - Labour productivity growth (lagged), ΔLP_{t-2} - Labour productivity growth (lagged 2 years), $\Delta InvSh_{t-1}$ - Change in the investment-to-GDP ratio (lagged), $\Delta YouthRatio_t$ - Change in the youth-to-total population share, $\Delta WPopSh_t$ - Change in the working age-to-total population share, $VulnEmpSh_{t-1}^{non-salaried}$ - Share of vulnerable workers in labour force minus salaried employment (lagged), $NonWorkPoorSh_{t-1}^{\$2}$ - Share of non-working poor using the \$2-a person-a day threshold (lagged), $\Delta VulnEmpSh_{t-1}^{non-salaried}$ - Change in the share of vulnerable workers in labour force minus salaried employment (lagged), $\Delta NonWorkPoorSh_{t-1}^{\$2}$ - Change in the share of non-working poor using the \$2-a person-a day threshold (lagged), $WorkPoorSh_{t-1}^{\$2}$ - Share of working poor using the \$2-a person-a day threshold (lagged), $I_t^{WorkPov \times VulnEmpSh^{non-salaried}}$ - Interaction dummy between working poor and share of vulnerable employment in labour force minus salaried employment, $WorkPoorRatio_{t-1}^{\$2}$ - Share of working poor in total poverty share using the \$2-a person-a day threshold (lagged), $I_t^{WorkPovSh \times VulnEmp^{non-salaried}}$ - Interaction dummy between working poor in poverty share and vulnerable employment in labour force minus salaried employment, $DepRatio_{t-1}$ - Dependency ratio (lagged), $I^{Crisis} \times \sum_i IndepVar_i$ - Crisis interaction dummy (with all incl. variables) , $Year \times \sum_i IndepVar_i$ - Year interaction dummy (with all incl. variables), $\Delta EmpSectorSh_{t-1}^{Broad}$ - Change in sectoral employment shares (broad sector, lagged), $EmpSectorSh_{t-1}^{Broad}$ - Sectoral employment share (broad sector, lagged), $\Delta EmpSectorSh_{t-1}^{Detail_{14}}$ - Change in sectoral employment share (14 disaggregated sectors, lagged), $\Delta EmpSectorSh_{t-1}^{Detail_4}$ - Change in sectoral employment share (4 disaggregated sectors, lagged), $EmpSh_{t-1}^{Occupation-by-task}$ - Occupational employment share (by tasks, lagged), $\Delta EmpSh_{t-1}^{Occupation-by-task}$ - Change in occupational employment share (by tasks, lagged), $EmpSh_{t-1}^{Occupation-by-skill}$ - Occupational employment share (by skills, lagged), $\Delta EmpSh_{t-1}^{Occupation-by-skill}$ - Change in occupational employment share (by skills, lagged), I^{Crisis} - Crisis dummy, $I^{Recovery}$ - Recovery dummy.

Table 3: Specification overview: Difference between real wages and productivity

Dependent variable: $\Delta w_t - \Delta LP_t$	1	2	3	4	5	6	7	8	9
ΔLP_{t-2}	■	■	■	■	■	■	■	■	■
$\Delta InvSh_{t-1}$	■	■	■	■	■	■	■	■	■
$\Delta WPopSh_t$	■	■	■	■	■	■	■	■	■
$\Delta EmpSectorSh_{t-1}^{Broad}$		■							
$EmpSectorSh_{t-1}^{Broad}$			■						
$\Delta EmpSectorSh_{t-1}^{Detail_{14}}$				■					
$\Delta EmpSectorSh_{t-1}^{Detail_{4}}$					■				
$EmpSh_{t-1}^{Occupation-by-task}$						■			
$\Delta EmpSh_{t-1}^{Occupation-by-task}$							■		
$EmpSh_{t-1}^{Occupation-by-skill}$								■	
$\Delta EmpSh_{t-1}^{Occupation-by-skill}$									■

Note: Δw_t - Real wage growth, ΔLP_t - Labour productivity growth, ΔLP_{t-2} - Labour productivity growth (lagged 2 years), $\Delta InvSh_{t-1}$ - Change in the investment-to-GDP ratio (lagged), $\Delta WPopSh_t$ - Change in the working age-to-total population share, $\Delta EmpSectorSh_{t-1}^{Broad}$ - Change in sectoral employment shares (broad sector, lagged), $EmpSectorSh_{t-1}^{Broad}$ - Sectoral employment share (broad sector, lagged), $\Delta EmpSectorSh_{t-1}^{Detail_{14}}$ - Change in sectoral employment share (14 disaggregated sectors, lagged), $\Delta EmpSectorSh_{t-1}^{Detail_{4}}$ - Change in sectoral employment share (4 disaggregated sectors, lagged), $EmpSh_{t-1}^{Occupation-by-task}$ - Occupational employment share (by tasks, lagged), $\Delta EmpSh_{t-1}^{Occupation-by-task}$ - Change in occupational employment share (by tasks, lagged), $EmpSh_{t-1}^{Occupation-by-skill}$ - Occupational employment share (by skills, lagged), $\Delta EmpSh_{t-1}^{Occupation-by-skill}$ - Change in occupational employment share (by skills, lagged).

Table 4: Specification overview: Log of real wages

Dependent variable: $\ln w_t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\ln LP_{t-1}$	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$\Delta InvSh_{t-1}$	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$\Delta WPopSh_t$	■	■	■	■	■	■	■	■	■	■	■	■	■	■
$VulnEmpSh_{t-1}^{non-salaried}$		■	■	■	■	■	■	■	■					
$NonWorkPoorSh_{t-1}^{\$2}$		■				■	■		■					
$WorkPoorSh_{t-1}^{\$2}$			■					■						
$I_t^{WorkPov \times VulnEmpSh^{non-salaried}}$			■											
$WorkPoorRatio_{t-1}^{\$2}$				■	■									
$I_t^{WorkPovRatio \times VulnEmpSh^{non-salaried}}$				■	■									
$DepRatio_{t-1}$					■									
$I^{Crisis} \times \sum_i IndepVar_i$							■	■						
$Year \times \sum_i IndepVar_i$									■					
$EmpSectorSh_{t-1}^{Broad}$										■				
$EmpSectorSh_{t-1}^{Detail_1 4}$											■			
$\Delta EmpSectorSh_{t-1}^{Detail_4}$												■		
$EmpSh_{t-1}^{Occupation-by-task}$													■	
$EmpSh_{t-1}^{Occupation-by-skill}$														■
$I^{Recovery}$						■								

Note: $\ln w_t$ - Log of real wage, ΔLP_{t-1} - Labour productivity growth (lagged), $\Delta InvSh_{t-1}$ - Change in the investment-to-GDP ratio (lagged), $\Delta WPopSh_t$ - Change in the working age-to-total population share, $VulnEmpSh_{t-1}^{non-salaried}$ - Share of vulnerable workers in labour force minus salaried employment (lagged), $NonWorkPoorSh_{t-1}^{\$2}$ - Share of non-working poor using the \$2-a person-a day threshold (lagged), $WorkPoorSh_{t-1}^{\$2}$ - Share of working poor using the \$2-a person-a day threshold (lagged), $I_t^{WorkPov \times VulnEmpSh^{non-salaried}}$ - Interaction dummy between working poor and share of vulnerable employment in labour force minus salaried employment, $WorkPoorRatio_{t-1}^{\$2}$ - Share of working poor in total poverty share using the \$2-a person-a day threshold (lagged), $I_t^{WorkPovRatio \times VulnEmp^{non-salaried}}$ - Interaction dummy between working poor in total poverty ratio and vulnerable employment in labour force minus salaried employment, $DepRatio_{t-1}$ - Dependency ratio (lagged), $I^{Crisis} \times \sum_i IndepVar_i$ - Crisis interaction dummy (with all incl. variables) , $Year \times \sum_i IndepVar_i$ - Year interaction dummy (with all incl. variables), $EmpSectorSh_{t-1}^{Broad}$ - Sectoral employment share (broad sector, lagged), $EmpSectorSh_{t-1}^{Detail}14$ - Sectoral employment share (14 disaggregated sectors, lagged), $\Delta EmpSectorSh_{t-1}^{Detail_4}$ - Change in sectoral employment share (4 disaggregated sectors, lagged), $EmpSh_{t-1}^{Occupation-by-task}$ - Occupational employment share (by tasks, lagged), $EmpSh_{t-1}^{Occupation-by-skill}$ - Occupational employment share (by skills, lagged), $I^{Recovery}$ - Recovery dummy.

5 Wage imputations and projections

5.1 Regression Results

The base specification, as accounted for by our theoretical model, analyses the relationship between wages and productivity (using past productivity and the investment-to-GDP ratio to account for expected productivity), and the expected outside option (using past vulnerable employment and non-working poverty rates), while controlling for demographic factors. Using this model as our starting point, we build up all our different specifications.

In addition to running different methodologies and dependent variables, we also look at other factors that could affect the wage level, such as the sectoral employment distribution, occupational employment and employment by skill level. In advanced economies, there are often stronger social security systems in place, therefore the outside option is often fixed. This means that the wage levels in the economy can be impacted to a greater extent through growth of specific sectors or industries, requiring specific skill sets. Table 5 shows the results of running a specification using different regression methodologies on employment by sector. Column 1 is a FE regression with clustered standard errors, column 2 is a GLS model, and column 3 is a ME model. Changing the estimation methodology can occasionally change our results, for example, by changing the significance of specific sectors. This allows us to capture different effects. Table 5 shows the results of an empirical model based on a sectoral model (accounting for the impact of different sectors on the wage level), and to ensure there is no collinearity in our independent regressors, we include only some of the 14 sectors that had data available. As expected, when productivity is significant, it has a positive impact on wage growth: as the economy becomes more productive, labour also becomes more productive, and wages start to rise. In this case, the results are provided for the entire panel, with no groupings.

5.2 Imputations

A significant part of our model is generating imputations for missing past data. The purpose of this is to ensure that we have a full set of past data in levels as well as in growth rates, which gives a balanced panel dataset for 112 countries for a period of nearly 25 years. This process is combined with the forecasting process, so that the series selected to fill in for a country is selected on a country-specific level. Additionally, since we use a variety of different dependent variables, this imputation at the wage level allows comparability across all different regression model outputs.

In order to calculate the wage levels, we use the growth rates predicted in section 5.1, and use the most recently available actual data point to backcast the series from the growth rate to calculate the level. For historic data for which recent wage data is missing, we use the predicted growth rates to forecast the wage level, and repeat this process for the forecasting years (after 2014). The newly calculated level is then used to calculate the level of the previous year, or alternatively the level of the next year, and this is repeated until we have a balanced panel set for all the countries for which it is possible to produce predictions. We complete this process for every model we have selected from the previous section; this leaves us with a very large variety of imputations to test.

Table 5: Example of Estimation Results with different methodologies

	(1)	(2)	(3)
	$\Delta RealWage$	$\Delta RealWage$	$\Delta RealWage$
$\Delta Prod_{t-1}$	0.321**	0.287**	0.458**
	(3.05)	(7.72)	(4.60)
$\Delta(InvmtGDPRatio_{t-1})$	0.000995	0.000484	0.000618
	(1.39)	(1.12)	(0.81)
$\Delta(EmpAgriculture_{t-1})$	0.00201	-0.000303	0.00148
	(0.81)	(-0.21)	(0.61)
$\Delta(EmpManuf_{t-1})$	0.00368	-0.00127	0.00239
	(1.21)	(-0.82)	(0.75)
$\Delta(EmpConstruction_{t-1})$	0.00340	0.00674**	0.00709**
	(0.96)	(3.47)	(2.16)
$\Delta(EmpTransportStorage_{t-1})$	0.00433*	0.00300**	0.00568**
	(1.71)	(2.02)	(2.23)
$\Delta(EmpFinancialAct_{t-1})$	0.0112**	0.00726**	0.00988**
	(3.13)	(2.57)	(2.46)
Constant	-0.0535**	0.00384**	0.00752
	(-2.68)	(2.31)	(1.56)
N	1108	1108	1108

Note: t-statistics in parentheses. Asteriks indicate the level of p-values with * p <10 % , ** p <5 %.
Variables included are $\Delta Prod_{t-1}$: lagged productivity growth, $\Delta(InvmtGDPRatio_{t-1})$: lagged share of investment-to-GDP, $\Delta(EmpAgriculture_{t-1})$: lagged growth of agricultural employment, $\Delta(EmpManuf_{t-1})$: lagged growth of manufacturing employment, $\Delta(EmpConstruction_{t-1})$: lagged growth of construction employment, $\Delta(EmpTransportStorage_{t-1})$: lagged growth of employment in transport and storage services, $\Delta(EmpFinancialAct_{t-1})$: lagged growth of employment in financial activities.

5.3 Testing and Forecasting

Once we have imputed all our values, we can start testing our forecasts to select the best-performing model. This entire process is automated using a six-step pseudo out-of-sample testing. To run this test, we initially restrict our data series to 2005, and develop a forecast six years in to the 'future' (so until 2011). We then take the percentage difference between the actual and estimated data for every forecasted year, at the country level. We then repeat this process for 2006-2012, observing that for 2012 it would be possible to have at maximum only two steps forward. Collecting together all our forecast errors, we create a weighted average of the first-step average for every year. The weighting is based on the number of years of data available; therefore more recent years (for example the one-step forecast in 2012 for 2013) are given a higher weight in the average. We can then sort the models according to those with the smallest forecast error. We select the top 10 best-performing models based on these errors.

After selecting the top 10 best-performing models, we average the forecasts as well as the imputations for that specific model, which gives us a final series for each country from 1999-2019. Since the averaging is done at the country-level, each country has a different set of top 10 best-performing models, which makes sense as no two countries have an identical macroeconomic environment, hence there will be differences in which models better represent the wage growth in that country.

Choosing to use an average of the top 10 models is helpful in the sense that it makes our forecast more robust to small changes in the models. Therefore, in the event that a different model is tested or, more likely, the set of best-performing models changes from year-to-year; our overall forecast does not change dramatically from year-to-year, as the final forecast is an average. It seems highly unlikely that the full set of all 10 models would change from year to year. We also try taking the top-20 and top-30 averages, but there is minimal difference to the final result, and therefore we deemed it unnecessary to average across more than 10.

5.3.1 Final Results

After generating wage growth data at the country-level, we aggregate these data by region, following the methodology introduced by the ILO's Global Wage Report (see ILO, 2014, p. 73). When aggregating, wages are weighted by the output of wage and salaried workers in a country as a share of the regional output, assuming that labour productivity of wage and salaried employees is equal to the economy-wide labour productivity. Table 6 and figure 2 show the results of the global and regional wage and productivity growth estimates and projections, respectively.

The charts highlight the need for looking at relationships outside of the labour productivity-wages linkage. Looking at the forecasts produced by this model, we see that the relationship between wages and productivity is quite varied by region. In Developed Economies for example productivity growth is currently outstripping wage growth, whereas in the majority of the other regions, this relationship is reversed. This result is in line with what we would expect of these economies; many developing economies still depend heavily on labour intensive growth and development, whereas the advanced economies have a greater proportion of capital intensive growth, which limits wage growth. The US especially has recently been seeing a growing economy, with wage stagnation.

Table 6: Global wage and productivity averages

	2000-07	**2008-09**	**2010-13**	**2014-16**	**2017-19**
Wage growth	1.9	1.6	1.7	2.5	3.1
Productivity growth	2.2	-1.0	2.4	2.5	2.8

Source: Ernst (2015) on the basis of ILO, Wage projection model, 2015

Table 7 shows the average forecast errors by region for the top 10 best-performing models for one run of the pseudo out-of-sample (2012, first step). As can be seen from the table, all forecasts are less than a percentage point deviation away from the true value. As the amount of past information available decreases, the forecast error increases, hence Sub-Saharan Africa has a higher error percentage than Developed Economies for example.

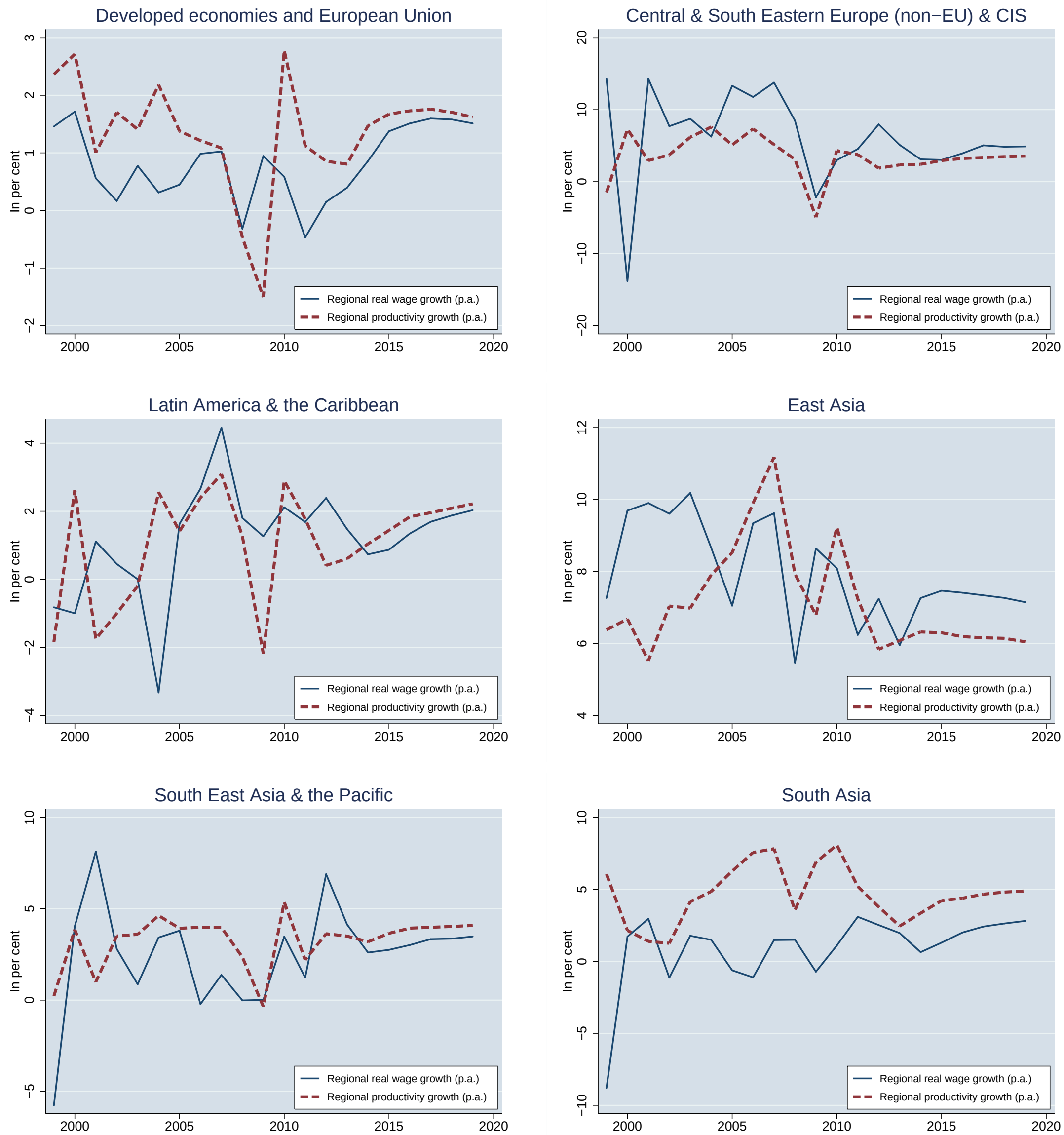
Table 7: Forecast Error by Region

Region	Average Forecast Error (%)
Central and South Eastern Europe and CIS	0.58
Developed Economies	0.55
East Asia	0.15
South-East Asia and the Pacific	0.16
South Asia	0.26
Latin America and the Caribbean	0.25
Middle East	0.30
North Africa	0.61
Sub-Saharan Africa	0.80

Note: The forecast errors refer to the specifications with real wage growth rates.

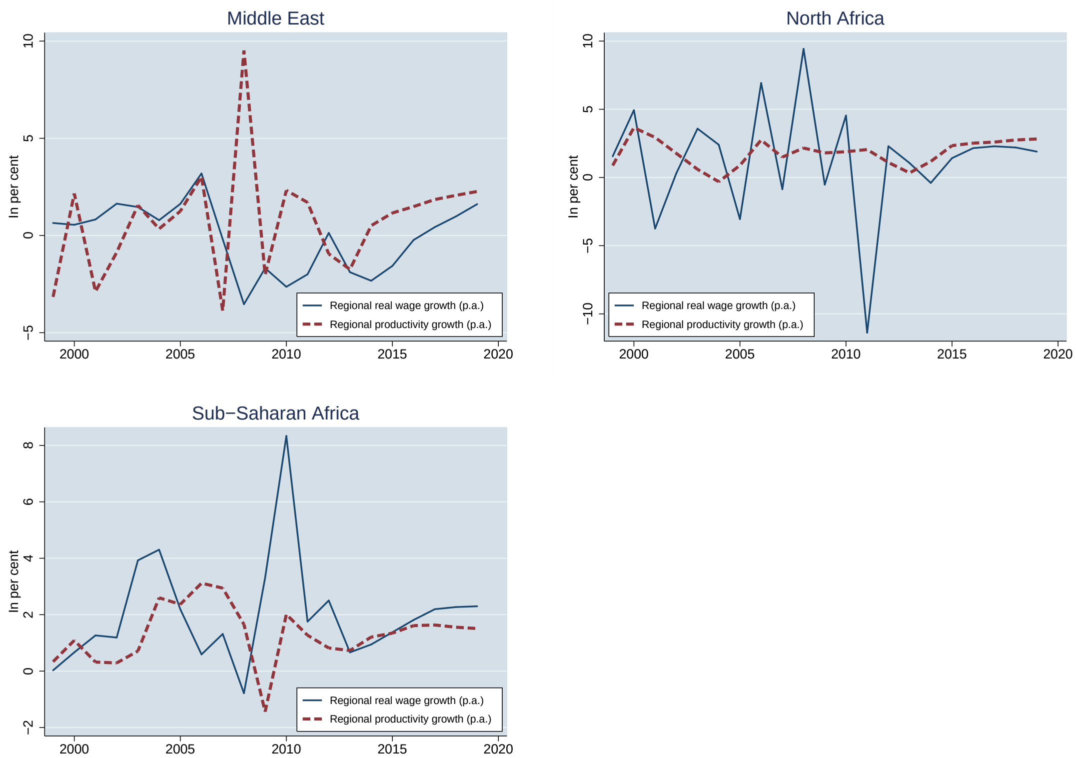
As the testing is carried out at the country level, each country has its own ranking for the models we look at, and therefore the forecasts generated are specific to the models that best fit the wage relationship of the country in question.

Figure 2: Regional wage and productivity developments



Source: ILO, Wage projection model, 2015

Figure 2: Regional wage and productivity developments (cont'd)

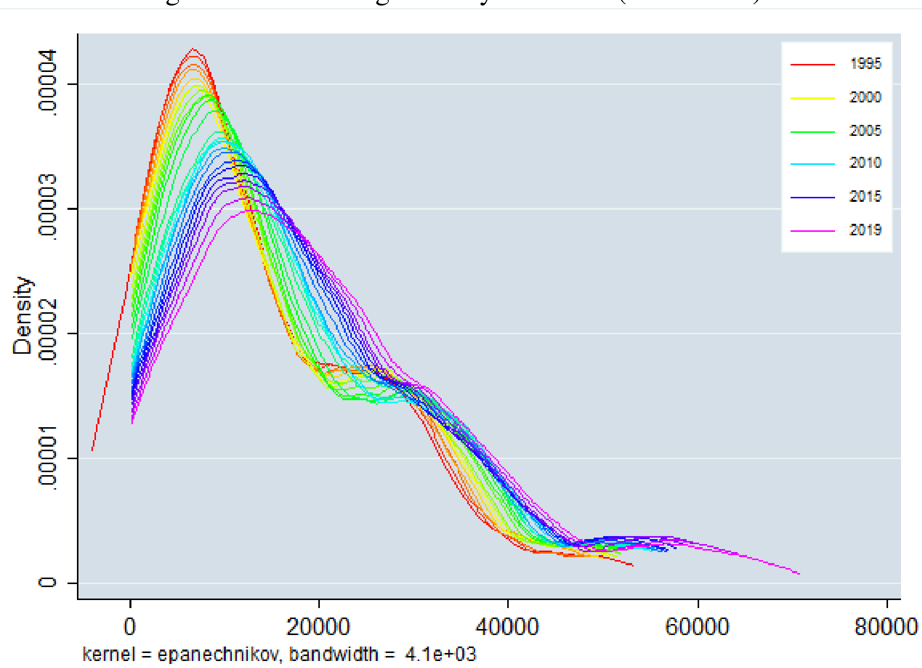


Source: ILO, Wage projection model, 2015

5.3.2 Wage developments

The development of a balanced panel with comparable wage information allows for a more detailed analysis of (expected) wage developments across countries and over time. In particular, it allows to better understand to what extent a (conditional) convergence of wage levels (similar to one that is expected to happen for per capita output levels) is taking place. The following chart plots that (kernel) distribution estimates at 5-years intervals of our wage estimations across the sample countries.

Figure 3: Kernel wage density estimates (1995-2019)



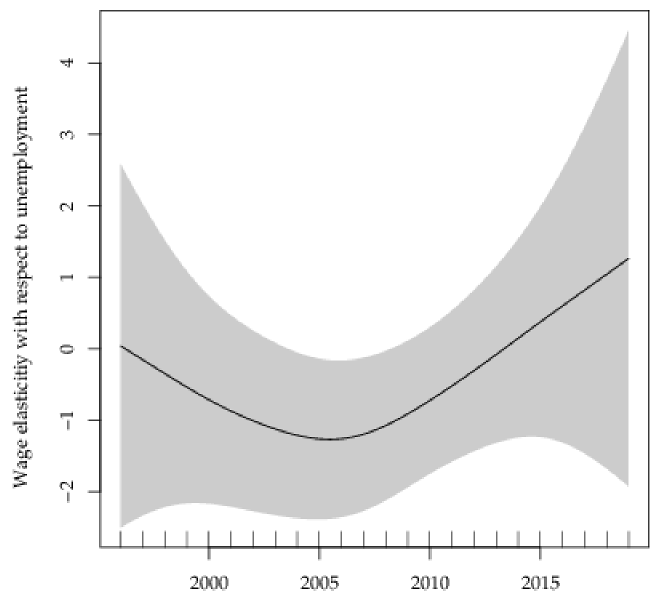
As the chart demonstrates, a gradual increase in wages has taken place (the mode of the distribution is moving to the left) but at the same time the variance of the distribution has increased. More importantly, a second mode seems to have developed at a significant higher level (making it a bi-modal distribution with "twin-peaks", a concept discussed by Durlauf and Johnson (1995) and Quah (1996)). This could potentially be an indication of a barrier to development for middle-income countries to rejoin high-income and -wage countries due to the existence of a middle-income trap.[7]

In addition to comparing trends across countries, our database also allows a long-term analysis of trends within specific regions. In particular, the database can be used to shed some light on recent

[7] See ILO (2015) for a discussion of the concept and potential underlying reasons that explain such a trap.

shifts in the wage Phillips curve (see figure 4). The figure demonstrates that the wage elasticity for the Developed Economies region as a whole has decreased in absolute values in recent years after having been significantly negative during the 2000s. This is in line with recent country-level evidence for the United States and the United Kingdom indicating that despite significant improvements in labour market conditions, wages have not yet started to accelerate. The figure indicates that this seems to be a region-wide issue and not only confined to specific countries.

Figure 4: Time-varying estimates of wage-unemployment elasticities - Developed Economies



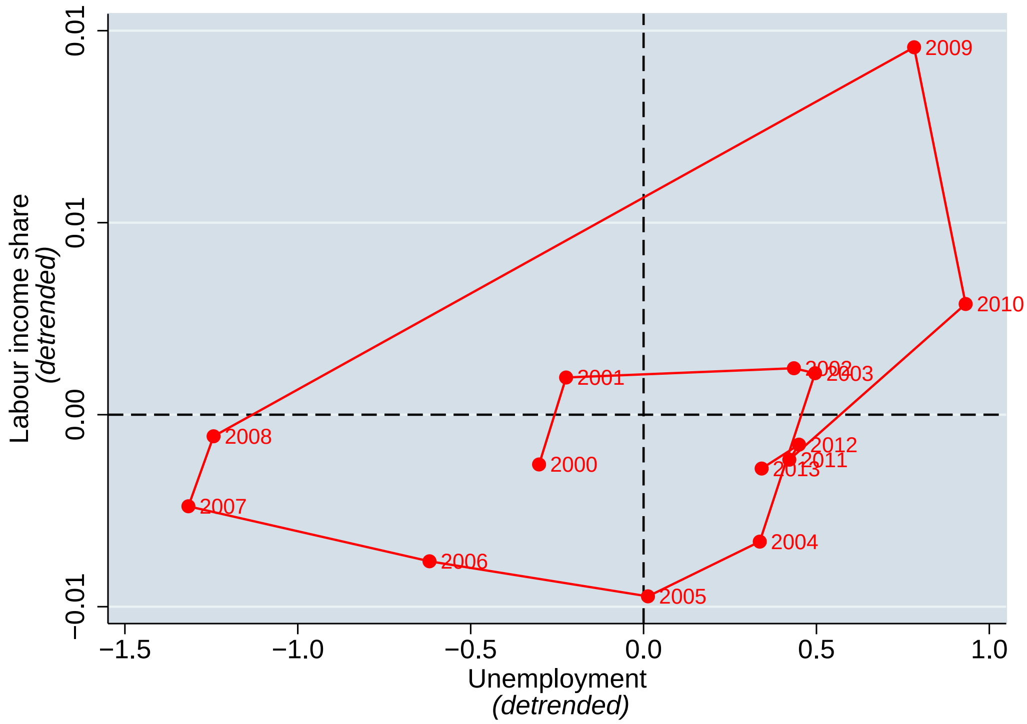
Note: The chart shows the time-varying coefficient, μ_t, of the elasticity of real wage growth with respect to changes in unemployment between 1994 and 2019, including forecast changes beyond 2014, for the Developed Economies and EU region. The shaded areas represent the confidence interval. Estimates have been established using P-splines with the estimation equation:

$$w_t = \beta^w \sum_{i=1}^{4} w_{t-i} + (\beta^u + \mu_t) \sum_{i=0}^{4} u_{t-i} + \varepsilon_t$$

where w_t: (regional) real wage growth, Δu_t: change in the (regional) unemployment rate and β^w, β^u: the (constant) elasticities of real wage growth with respect to past wage growth and changes in unemployment.
Source: ILO (2015)

Moreover, our wage estimations allow to take a closer look at longer term links between wages and unemployment (see figure 5 that depicts the so-called Goodwin cycle, referring to the original work by Goodwin, 1967). The chart demonstrates that over longer periods of time, the labour income share and unemployment rates follow a cyclical pattern, an indication of systematic over- and under-shooting in wage developments. Such wage patterns can arise when bargaining between employers and workers over the income distribution does not have a fully time-consistent horizon but aims at maximizing current distributional gains. As such it is consistent with hyperbolic discounting, finite time horizons and right-to-manage wage bargaining models.

Figure 5: The Goodwin growth cycle



Source: Own calculations.

6 Conclusion

The aim of this paper was to provide a methodology for forecasting wage growth, and imputing missing data points in historical data to produce a balanced panel dataset. This empirical model is based on two potential theoretical frameworks that were presented at the beginning of the paper. This gave us a solid foundation on which to build up our econometric model, highlighting which variables we needed to focus on. For our first model, labour productivity measures and measures of vulnerable employment were used to proxy for the relationship in wage bargaining, while for our second model we instead used employment by sector and/or occupation with labour productivity measures.

Using a variety of regression methodologies allows us to capture a huge variety of different effects of wages; our fixed effect model helps to control for non-time varying factors, whilst our mixed models help us to control for the idea that there may be correlations between wage growth between countries, especially given our construction of country groupings. This helps us develop a series of possible predictions for our model, which we then test using pseudo-out-of-sample testing. This process allows us to create an automatic mechanism for selecting the best performing models, by restricting our available data and seeing how our models perform in comparison to actual data points. The prevalence of average forecast errors consistently below 1% is a testament to the success of this methodology. In addition, allowing the testing to run at the country-level allows us to also capture specificities in individual economies that we would otherwise be unable to do so due to lack of sufficient data to run country-level regressions. Once we have selected our top 10 best-performing models, we take a simple average of the entire wage series, which gives us our final series for the country as a whole.

The results presented in this paper are the most recent available (with respect to the latest available wage data), and it is possible to update this model annually to produce an extra year's worth of forecasts, as well as improving the testing of the model by the inclusion of a another year worth of information. Furthermore, both the IMF WEO and the ILO GET input data are updated near quarterly, which opens the potential for more frequent updating. However wage data is only available annually, therefore a full update of the model would only be possible annually.

Our current methodology allows us to generate data and forecasts for which we have a wave non-response missing pattern. A potential future exercise to carry out would be to develop a procedure that deal with countries that have no data at all (unit non-response), so this is a potential future exercise we could carry out. In addition, the use of Bayesian modelling procedures in our model development has the potential to add valuable information to our modelling procedure, and would be an interesting addition to the model. Analysing different model specifications could also be a useful direction to take, and this would require studying alternative theoretical frameworks to use as a base for our empirical specifications. Finally, this model is currently aimed at looking at aggregate wage growth in the economy as a whole. As has been discussed throughout this paper, the inclusion of sectoral information is important for capturing certain effects. Similarly, the analysis of aggregate wage data may hide more interesting sectoral movements. This is a potential advancement of the model, given adequate sectoral wage data.

As has been mentioned earlier, the model presented in this paper is meant to provide us with a basis for understanding the movement of wages over a future time horizon of 6 years. It provides no causal explanations, and is not a deterministic model; rather it allows us to shape a relationship between wages and a series of other macroeconomic variables, which allows us to glean some insights into the likely evolution of wages over the next few years.

References

Autor, D.H.; Levy, F.; Murnane, R.J. 2003. "The skill content of recent technological change: An empirical exploration", in *The Quarterly Journal of Economics*, vol. 118, no. 4, pp. 1279–1333.

Binmore, K.; Rubinstein, A.; Wolinsky, A. 1986. "The Nash bargaining solution in economic modelling", in *The RAND Journal of Economics*, pp. 176–188.

Blanchflower, D.G.; Oswald, A.J. 1994. *The wage curve* (Cambridge, MA., MIT Press).

Calmfors, L.; Driffill, J. 1988. "Bargaining Structure, Corporatism and Macroeconomic Performance", in *Economic Policy*, vol. 9, pp. 14–61.

Charpe, M.; Kühn, S. 2015. "Demand and supply effects of bargaining power shocks", in *Economic Modelling*, vol. 51, pp. 21–32.

Dreger, C.; Reimers, H.E. 2010. *On the role of sectoral and national components in the wage bargaining process*, Discussion Paper 4908 (Bonn, Institute for the Study of Labor (IZA)).

Duncan, G.M.; Leigh, D.E. 1980. "Wage determination in the union and nonunion sectors: A sample selectivity approach", in *Industrial & Labor Relations Review*, vol. 34, no. 1, pp. 24–34.

Durlauf, S.N.; Johnson, P. 1995. "Multiple regimes and cross-country growth behavior", in *Journal of Applied Econometrics*, vol. 10, no. 4, pp. 365–384.

Ernst, E. 2015. "The shrinking middle.", in *Finance and Development*, vol. 52, no. 1, pp. 20–23.

Geary, R.C. 1950. "A note on 'A constant-utility index of the cost of living"', in *Review of Economic Studies*, vol. 18, no. 2, pp. 65–66.

Goodwin, R.M. 1967. "A growth cycle", in *Socialism, Capitalism and Economic Growth* (ed. C.H. Feinstein) (Cambridge, Cambridge University Press).

Hawksworth, J.; Lambe, C. 2013. *Global wage projections to 2030* (Lnodon, PriceWaterhouseCoopers).

ILO. 2014. *Global wage report. Wages and income inequality* (Geneva, International Labour Organization (ILO)).

—. 2015. *World employment and social outlook. Trends* (Geneva, International Labour Organization (ILO)).

Lee, K.C.; Pesaran, M.H. 1993. "The role of sectoral interactions in wage determination in the UK economy", in *The Economic Journal*, vol. 103, no. 416, pp. 21–55.

Levanon, G. et al. 2016. *Labor Cost Projections*, Economics Program Working Paper Series 1603 (New York, The Conference Board).

McDonald, I.M.; Solow, R.M. 1981. "Wage bargaining and employment", in *American Economic Reivew*, vol. 71, no. 5, pp. 896–908.

Quah, D.T. 1996. "Twin peaks: Growth and convergence in models of distribution dynamics", in *The Economic Journal*, vol. 106, no. 437, pp. 1045–1055.

Stone, R. 1954. "Linear expenditure systems and demand analysis: An application to the pattern of british demand", in *Economic Journal*, vol. 64, pp. 511–527.

Whelan, K. 2000. *Real wage dynamics and the Phillips curve*, Finance and Economics Discussion Series 2000-02 (Washington, DC, Federal Reserve Board).