Online Appendix: ARDL bounds test for cointegration: Replicating the *Pesaran et al. (2001)* results for the UK earnings equation using R

Kleanthis Natsiopoulos, Nickolaos G. Tzeremes
Department of Economics, University of Thessaly, 28th October street, 78, 38333, Volos, Greece

Abstract

This paper replicates the UK earnings equation using the autoregressive distributed lag (ARDL) modeling approach and the bounds test for cointegration by Pesaran et al.(*Journal of Applied Econometrics*, 2001, 16(3), 289–326). The findings from the narrow sense fully replicate the original results using the open source language R and the ARDL package. In the wide sense replication, augmented data are employed, thus extending the end period from 1997:Q4 to 2019:Q4, using an alternative measure for union power. Adopting the new dataset, this study reinvestigates the UK earnings equation, thereby providing supporting evidence of a long-run relationship and reveals empirical findings about the long-run effects of productivity, unemployment, tax wedge, and union power on wages.

Keywords: bounds test, cointegration, ARDL, earnings equation, R.

A. Introduction

This document is the Online Appendix included in the Supporting Information. It accompanies the main article of Natsiopoulos and Tzeremes (2021), "ARDL bounds test for cointegration: Replicating the Pesaran et al. (2001) results for the UK earnings equation using R", thus providing more details about the modeling process and the main results.

It performs a narrow and wide sense replication study of Pesaran, Shin, and Smith (2001) (henceforth PSS), "Bounds testing approaches to the analysis of level relationships", using the R language (R Core Team 2020)[1] and the **ARDL** package (Natsiopoulos and Tzeremes 2021).

The original paper by PSS is known for developing the widely used bounds test for cointegration. The main objective of the narrow sense replication section is to reproduce the results of the UK earnings equation. However, as open-source software comes with no guarantee, this fully reproducible replication study could serve as a means of validating the software calculations for these types of models and tests.

It must be highlighted that wide and narrow sense replications are strictly referred to in the original PSS methodological framework. Several other advances of the autoregressive distributed lag (ARDL) model, such as the nonlinear ARDL, threshold ARDL, spatiotemporal ARDL, quantile ARDL, and nonlinear quantile ARDL models have been developed. However,

[1] R version 3.6.3 (2020-02-29)

such methodological advances are beyond the scope of this study. For an insightful comparison and representation of the recent developments in ARDL modelling, see the literature review by Cho, Greenwood-Nimmo, and Shin (2021).

Multivariate analysis is another approach to derive cointegrated models. Several methodologies, such as the one developed by Johansen (1991), use vector autoregressive (VAR) and vector error correction models analogous to the use of the ARDL and error correction models (ECM) in the univariate conditional case presented here. Although the question under investigation is the same in both approaches, the existence of a cointegrated relationship, the underlying hypotheses, the model structure, and the results of the analysis may differ in practice. Through the multivariate analysis, more than one cointegrating vector may be found. This is, by design, explicitly restricted to at most one in the univariate conditional case. Similarly, the choice of the dependent variable is assumed a priori in the univariate analysis, while this restriction does not hold under a VAR model. The important properties of the multivariate approach may lead to richer findings through the dynamics between the data and possibly multiple error correction mechanisms. As each method has its own advantages, one should not conclude based on only one model but consider various alternatives. However, multivariate analysis is not used herein as it exceeds the scope of this replication study.

Section B describes the structure of the online Supporting Information. Next, in Section C, the two datasets used in the narrow and wide sense replications are described in detail, following every step of converting the raw data into the final datasets. Section D describes the basic equations that are used in the modeling and testing processes. In Section E, the results of PSS are successfully reproduced using the exact same dataset. In Section F, the period under investigation is extended using updated and augmented data to reinvestigate the UK earnings equation. Finally, in Section G, a robustness analysis is applied to the wide sense replication model to better understand the model dynamics and the UK economy regarding the earnings equation.

B. Reproducibility

The analysis was performed using the R language (R Core Team 2020) and the **ARDL** package (Natsiopoulos and Tzeremes 2021). Other packages such as **aTSA**, **CADFtest**, **dynlm**, **ggfortify**, **ggplot2**, **ggpubr**, **lmtest**, **olsrr**, **seasonal**, **strucchange**, **tseries**, and **xtable** (Qiu 2015; Lupi 2009; Zeileis 2019; Tang, Horikoshi, and Li 2016; Wickham 2016; Kassambara 2020; Zeileis and Hothorn 2002; Hebbali 2020; Sax and Eddelbuettel 2018; Zeileis, Leisch, Hornik, and Kleiber 2002; Trapletti and Hornik 2019; Dahl, Scott, Roosen, Magnusson, and Swinton 2019) were also used to preprocess the data and support the modeling, testing, and presentation of results.

This replication is the closest to the gold standard for a full replication, as it provides linked and executable codes and data (Peng 2011). The main files provided in the online Supporting Information are the codes that fully reproduce the entire replication procedure, the data used in the narrow and wide sense replication, the figures, and other files that are used to compile the present Appendix.

The structure of the online Supporting Information is described below. This contains the files `data_preparation_narrow.R` and `data_preparation_wide.R`, which process the raw data and produce the final datasets; the files `narrow_replication.R` and `wide_replication.R`, which comprise all the calculations and details regarding the narrow and wide sense replication processes; and the file `robustness_analysis.R`, which contains the code for the relevant section. It also includes the file `functions.R`, wherein some custom functions that are used in the previously mentioned files are defined.

The following folders are also included:

- `data`
- `figures`
- `appendix`

The figures used in the main paper and in the present Appendix are in the folder `figures`, while the folder `appendix` contains the files `appendix.rnw`, `jss.cls`, and `reph.bib` to produce the Appendix directly using R, and some other files needed to build the `appendix.pdf` file from the `appendix.tex` file.

The folder `data` includes the following files: for file names consisting of two words separated by "-", the first word represents the variable name and the second denotes the Office for National Statistics (henceforth ONS) identifier:

- `earn1.dat`
- `earn2.dat`
- `data_clean_narrow.csv`
- `data_clean_wide.csv`
- `AIIH-AIIH.csv`
- `ECG-G6NQ.csv`
- `ELA-G6NT.csv`
- `EMPNIC-CEAN.csv`
- `EMPSC-ROYK.csv`
- `GVA-ABMM.csv`
- `LFSUR-MGSX.csv`
- `NIS-GTAY.csv`
- `PRXMIP-CHMK.csv`
- `PVGA-CGBV.csv`
- `TYEM-DBBQ.csv`
- `TradeUnionMembers.csv`

- `WFJ-DYDC.csv`
- `WFP-DTWL.csv`
- `WRGTP-LOJU.csv`

The data files `earn1.dat` and `earn2.dat` contain the original data used in Pesaran *et al.* (2001)[2], which are employed for the narrow replication.

The files `data_clean_narrow.csv` and `data_clean_wide.csv` contain a clean form of the data that are used in the narrow and wide replication, respectively. These files are created by running the code in `data_preparation_narrow.R` and `data_preparation_wide.R`, respectively, using the unprocessed data in the files `earn1.dat` and `earn2.dat` for the narrow replication and the rest of the files in the above-stated list for wide replication.

The data used for the wide replication were extracted from the ONS database[3], with the exceptions of $UDEN$ and $UnionMembers$[4].

C. Data

The data used in PSS are also utilized for the narrow sense replication, covering the period 1970:Q1 to 1997:Q4, while the data for the wide sense replication cover the extended period from 1971:Q1 to 2019:Q4. Notably, PSS mention that they keep the first eight observations to construct the first differences and lagged variables; thus, all the regressions are calculated over the same sample period.

For the wide replication data, following the estimation strategy of the narrow sense replication, the first seven observations for the period 1971:Q1–1972:Q3 are used as a battery, and every calculation is performed on the sample for the period 1972:Q4–2019:Q4.

Tables 1, 2, and 3 show a detailed description of the data used in both the narrow and wide replications, as well as their components, and further decomposition.

[2] Retrieved from the JAE Data Archive `http://qed.econ.queensu.ca/jae/2001-v16.3/pesaran-shin-smith/`

[3] `https://www.ons.gov.uk/`

[4] The primary source of $UDEN$ is the Department of Employment Gazette and of $UnionMembers$ is the Department of Employment Statistics Division (for the data from 1892–1974) and Certification Office (for the years 1974–2019)

Table 1: Main variables

Variable name	Variable description	Narrow variable composition	Wide variable composition
w	real wage	$ln\left(\frac{ERPR}{PYNONG}\right)$	$ln\left(\frac{ERPR}{PGVA}\right)$
Prod	labor productivity	$ln\left(\frac{YPROM+278.29*YMF}{EMF+ENMF}\right)$	$ln\left(\frac{GVA}{EPS}\right)$
UR	unemployment rate	$ln\left(\frac{100*ILOU}{ILOU+WFEMP}\right)$	$ln(LFSUR)$
Wedge	wedge effect	$ln\left(\frac{(1+TE)*(1-TD)*PYNONG}{RPIX}\right)$	$ln\left(\frac{ERSTR*EESTR*PGVA}{PRXMIP}\right)$
Union	union power	$ln(UDEN)$	$ln(UDEN)$
D7475	income policies	$1:1974:Q1-1975:Q4$ $0:$ elsewhere	$1:1974:Q1-1975:Q4$ $0:$ elsewhere
D7579	income policies	$1:1975:Q1-1979:Q4$ $0:$ elsewhere	$1:1975:Q1-1979:Q4$ $0:$ elsewhere
UnionR	union membership rate	$-$	$ln\left(\frac{UnionMembers}{EPS}\right)$

Notes. Variable description refers to the quantity inside the ln (where applicable).
The variable $LFSUR$ have been used as a whole but it is further explained exactly like its PSS counterpart $LFSUR = \frac{100*ULFS}{ULFS+ETLFS}$.
$UnionR$ is an alternative measure for $Union$ and it is used only in the wide sense replication.

Table 2: Component variables for narrow replication

Variable name	Variable description	Variable composition
ERPR	average private sector earnings per employee (£)	–
PYNONG	the non-oil non-government GDP deflator	–
YPROM	output in the private, non-oil, non-manufacturing and public traded sectors at constant factor cost (£million, 1990)	–
YMF	manufacturing output index adjusted for stock changes (1990=100)	–
EMF	employment in UK manufacturing sectors (thousands)	–
ENMF	employment in UK non-manufacturing sectors (thousands)	–
ILOU	International Labor Office measure of unemployment (thousands)	–
WFEMP	total employment (thousands)	–
TE	the average employers' National Insurance contribution rate	$\frac{EMPNIC+NIS+OCR}{WFP}$
TD	the average direct tax rate on employment incomes	$\frac{TYEM+0.914*EENIC}{WFP}$
RPIX	the Retail Price Index excluding mortgage payments	–
UDEN	union density measured as union membership as a percentage of employment (constant from 1980:Q4)	–
EMPNIC	employers' payments of National Insurance Contributions (£million)	–
NIS	national insurance surcharge accruals (£million)	–
OCR	employers' other contributions (£million)	–
WFP	wage and salary bill (including forces' pay) (£million)	–
TYEM	accruals of tax on employment income (including PAYE accruing on retirement)	–
EENIC	employees' payments of National Insurance Contributions (£million)	–

Note. PYNONG is explained further in the `Readme file.txt` file in the online JAE Data Archive of PSS.

Table 3: Component variables for wide replication

Variable name	Variable description	Variable composition
ERPR	average private sector earnings per employee (£)	$\frac{WFP}{EPS}$
WFP	wages and salaries (incl. benefits in kind) (£million)	–
EPS	private sector employment (thousands)	WFJ−WRGTP−ECG−ELA
PGVA	gross value added deflator	–
GVA	gross value added at constant basic prices (£million)	–
WFJ	workforce jobs (thousands)	–
WRGTP	government trainees (thousands)	–
ECG	public sector employment central government (thousands)	–
ELA	public sector employment local government (thousands)	–
LFSUR	unemployment rate	–
ERSTR	employers tax rate	$1 + \frac{EMPSC+NIS}{WFP}$
EESTR	employees tax rate	$1 - \frac{TYEM+0.914*EENIC}{WFP}$
PRXMIP	RPI excluding mortgage interest payments	–
EMPSC	employers' social contributions (£million)	–
NIS	employers' national insurance surcharge (£million)	–
TYEM	taxes on income from employment (£million)	–
EENIC	employees' and self employed payments of NICs (£million)	AIIH−EMPNIC
AIIH	compulsory payments (£million)	–
EMPNIC	employers' payments of NICs (£million)	–
UDEN	rate of unionization in the labor market	–
UnionMembers	trade union members (thousands)	–

C.1. Data Preparation

Note that not all the variables listed in Table 3 are complete series or ready to be used to form the final variables listed in Table 1. In this subsection, the data preparation process for each variable component is described analytically. Each step that is described here is performed in the file `data_preparation_wide.R`, thus shaping the final form of the data.

The variables are already seasonally adjusted (where needed) or an adjustment is performed in the data preparation process as described below, while for the variable compositions mentioned in Tables 1 and 3, the component variables are used as described below.

ERPR: The final ratio is multiplied by 1000 as EPS is measured in thousands.

EPS: There are limited data, starting from 1999:Q1, to form the series, as described in Table 3. Therefore, these data are merged with the corresponding data from PSS (EMF+ENMF), which are used to describe the same variable. Thus, the variable has the appropriate time span (1971:Q1-2019:Q1), but there are still some missing data from 1998:Q1-1998:Q4, which are imputed via cubic spline interpolation.

LFSUR: This variable starts from 1971:Q1. The missing data (1970:Q1–1970:Q4) are not filled in using the values from the PSS data, as a suspicious drop is observed between 1969:Q1 and 1972:Q2. Hence, the variable UR starts from 1971:Q1 instead of 1970:Q1, and so does the entire wide replication.

PRXMIP: It starts from 1987:Q1 and the past values are filled in using the PSS PRIX variable.

NIS: It starts from 1987:Q1 (while being all zeros) and the past values are filled in using the PSS NIS variable.

TYEM: It starts from 1987:Q1 and the past values are filled in using the PSS TYEM variable.

EENIC: The quantity (AIIH-EMPNIC) is properly constructed and it extends back to 1987:Q1. The rest of the past values are filled in using the PSS EENIC variable.

AIIH: A seasonal adjustment is performed on the variable.

EMPNIC: This variable is initially yearly based. The value for each year is shared equally among its quarters.

UDEN: For this variable, the same variable as in PSS is used until 1980:Q4, and it is kept constant[5]. The same variable is used in Office for Budget Responsibility (OBR) (2013) (henceforth OBR), as taken from the Department of Employment Gazette[6].

Figures 1 to 4 depict the variables used in the narrow sense replication, and Figures 5 to 10 illustrate those used in the wide sense replication. Herein, they are presented separately, providing more details than the grouped figures in the main article.

[5] Strictly speaking, a very small change (by 0.00123 in log scale is detected) from 1993:Q1 onward and in the original PSS data.

[6] PSS mention that it measures the union membership as a percentage of employment, but it is more accurate to refer to this as OBR describes it, viz., "a proxy for structural changes in the labor market"

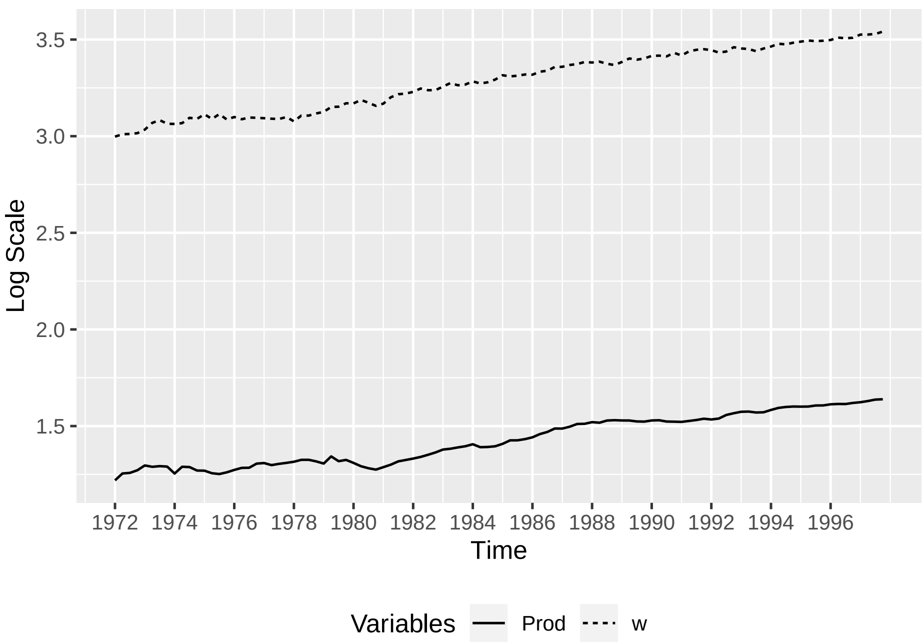

Figure 1: Real wages and labor productivity (Figure 1(a) in PSS)

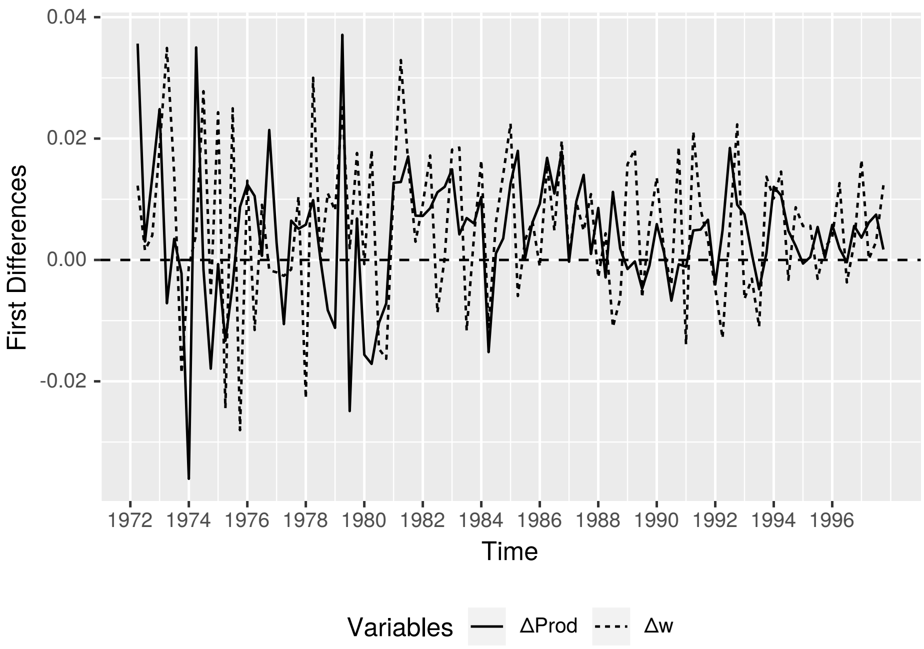

Figure 2: Rate of change of real wages and labor productivity (Figure 1(b) in PSS)

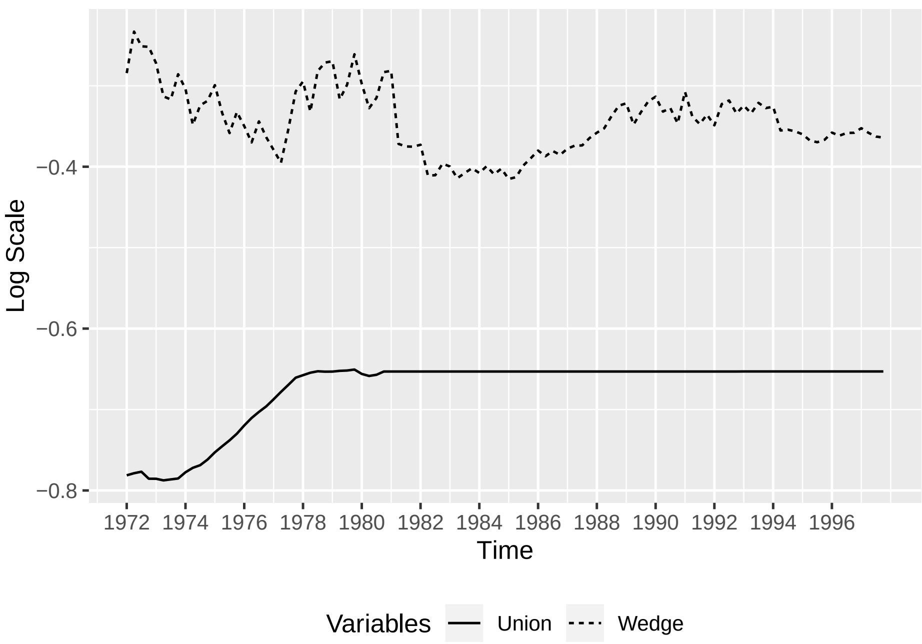

Figure 3: The Wedge and the unionization variables (Figure 2 in PSS)

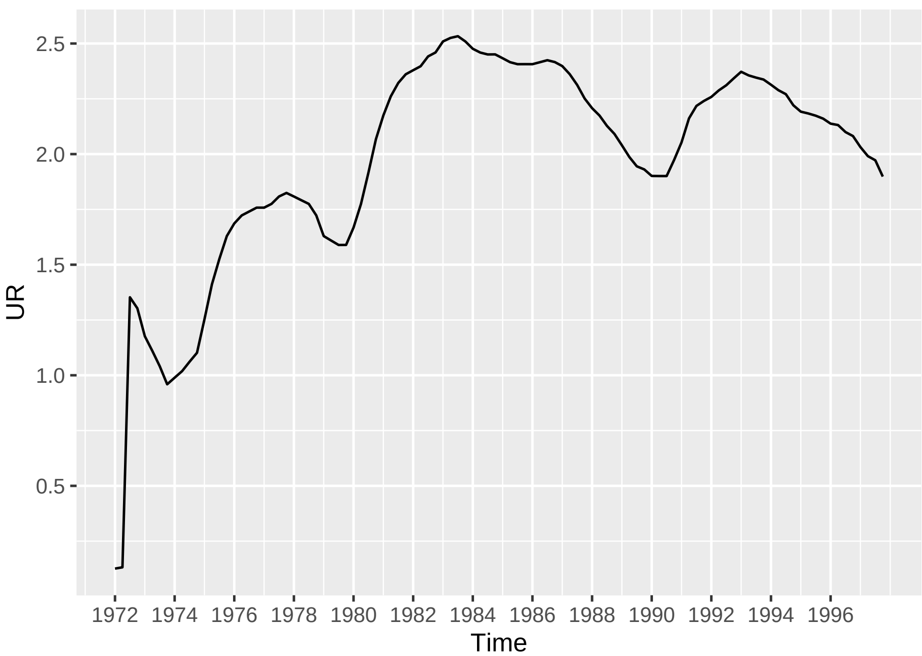

Figure 4: The unemployment rate (Figure 3 in PSS)

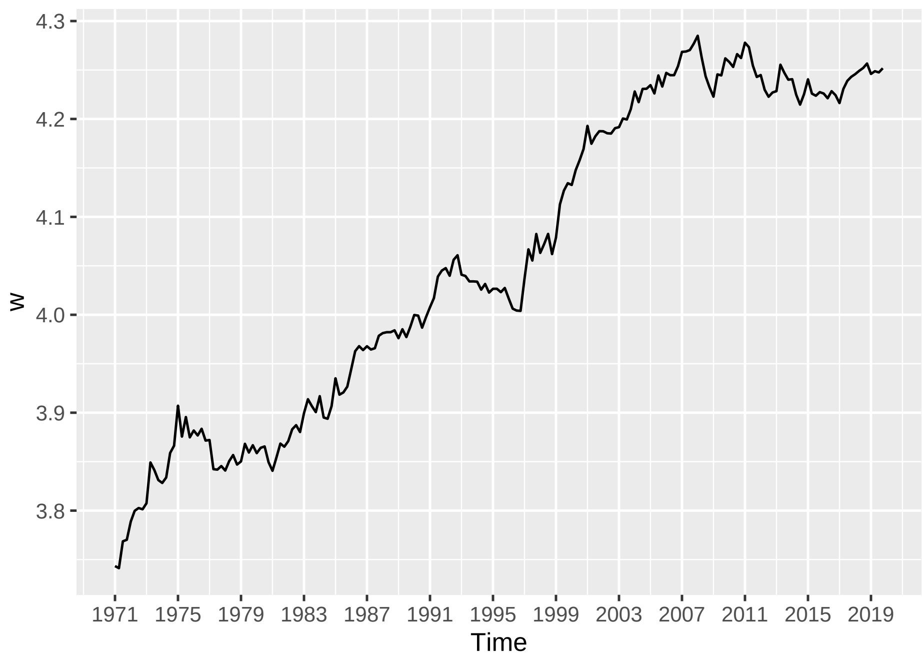

Figure 5: Real wages (w)

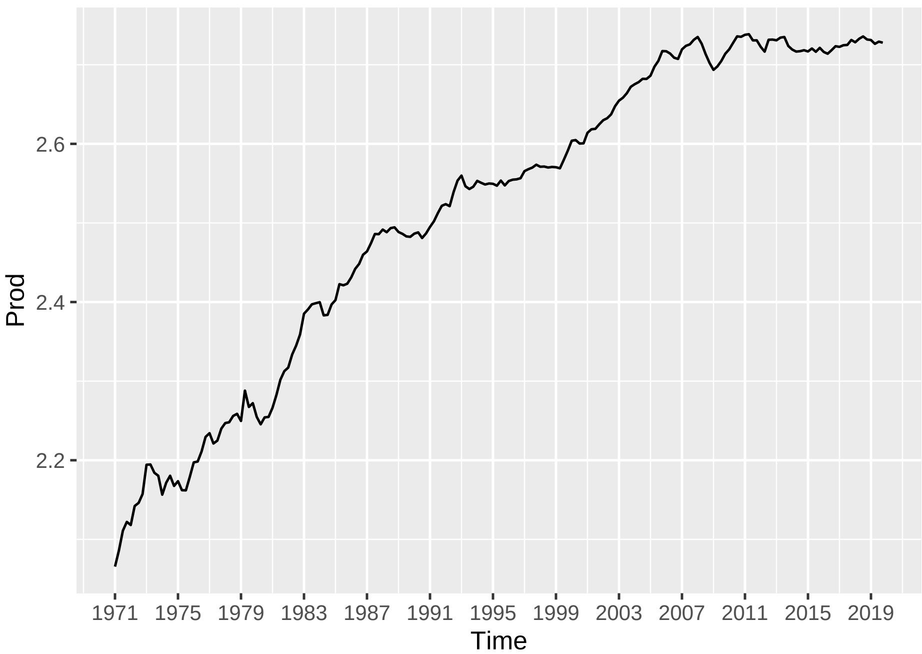

Figure 6: Labor productivity (Prod)

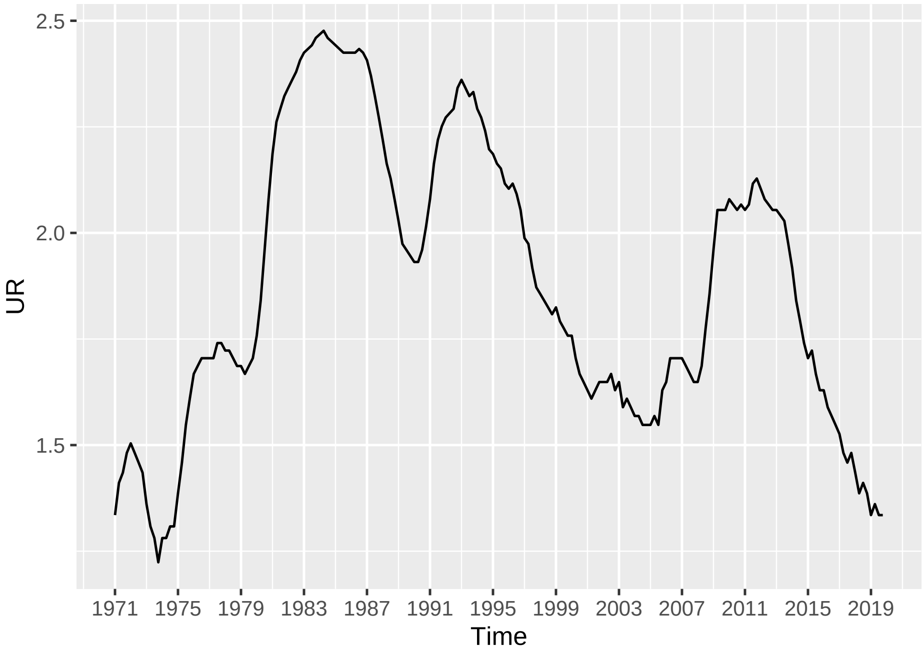

Figure 7: Unemployment rate (UR)

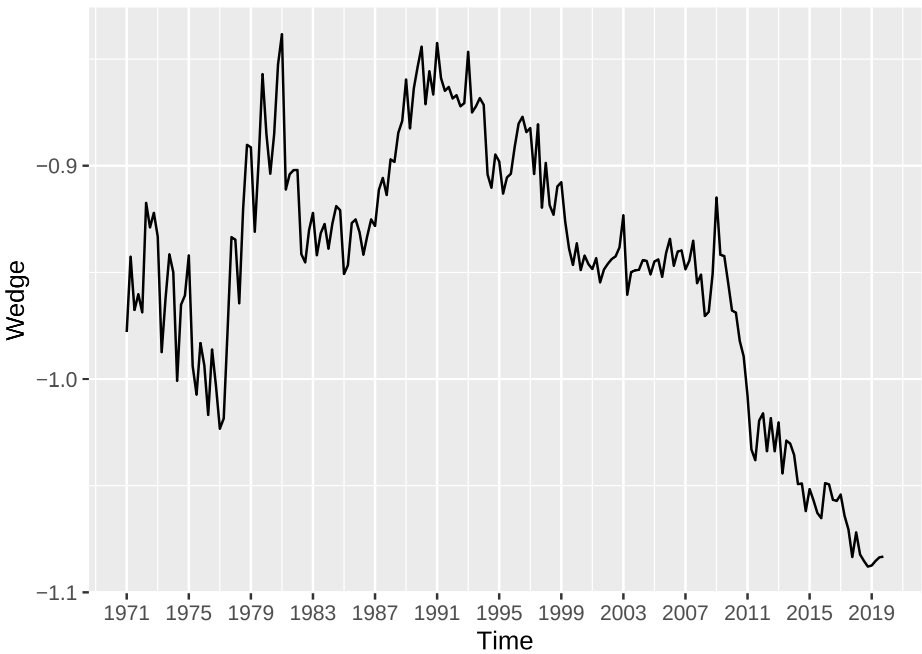

Figure 8: Wedge effect (Wedge)

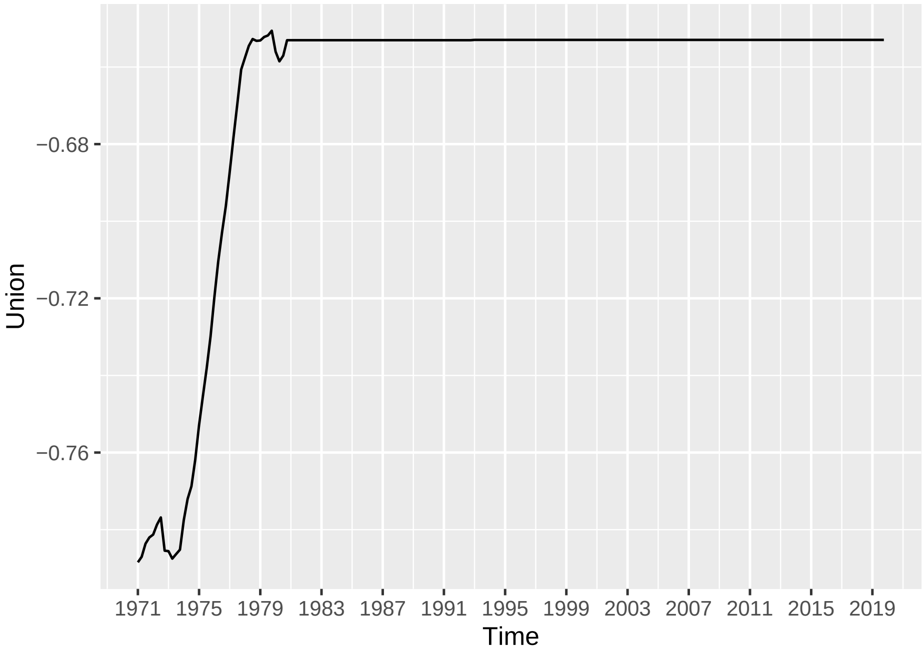

Figure 9: Unionization (Union)

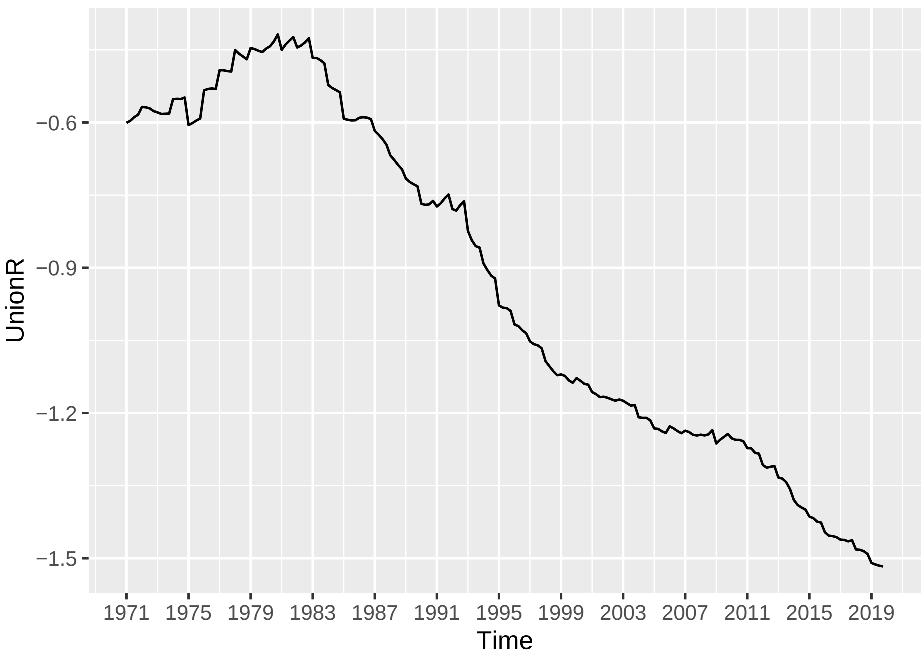

Figure 10: Trade union membership rate (UnionR)

C.2. Differences Between Narrow and Wide Data

The results from the wide sense replication, as explained in the following sections, differ from those of the narrow sense replication not only because of the extended period but also because of some differences in the data[7].

Most of the variables are identical at the shared period. The variable that is different in terms of definition, while it describes the same concept, is the dependent variable w (real wage). As shown in Table 3, wages and salaries (WFP) are divided by the number of employees (EPS) instead of using the average weekly earnings, as these two approaches are conceptually similar. In contrast to the narrow sense model, there are practical reasons for this change. This is because the series of average weekly earnings stops back in 2000, despite being the official measure since 2010[8].

The second difference is about the denominator (deflator) of the real wage. For the narrow (wide) sense replication, PYNONG (PGVA) is used. This choice is also followed by OBR, and this approach is adopted for data availability reasons. The two deflators are presented in levels and log scale in Figures 11 and 12, respectively.

The choice of the deflator affects the real wages (w) and the variable $Wedge$, as it is a component of both. See Figures 16 and 17 in Section G for the comparison of real wages and $Wedge$ using the two deflators.

Figure 11: Deflators in levels

[7] see also Section G

[8] see Office for Budget Responsibility (OBR) (2013), where they estimate the same model

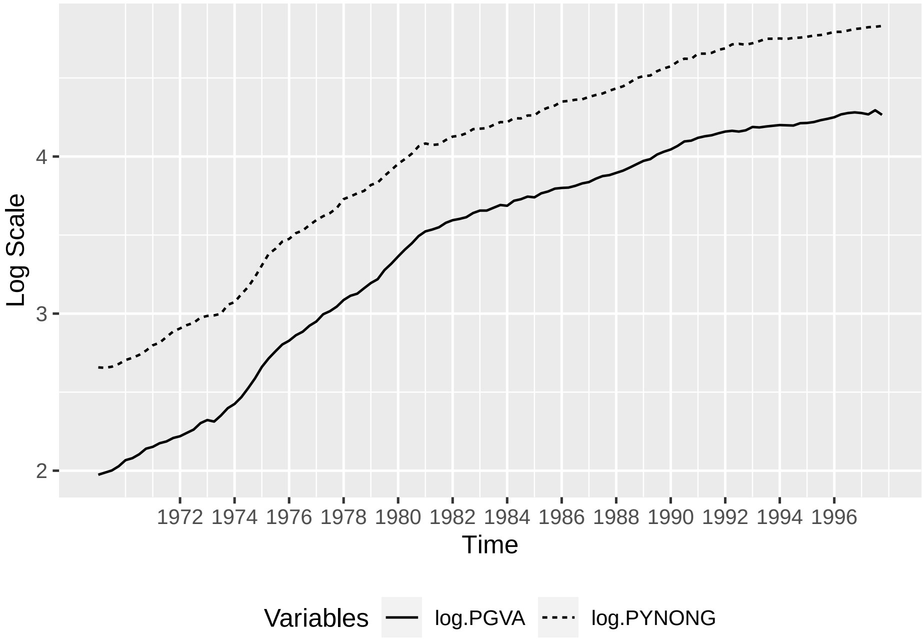

Figure 12: Deflators in log scale

D. Methodology

Both the narrow and wide sense replication studies examine the UK Treasury's earnings equation, which is based on the bargaining model (Nickell and Andrews 1983). As in the study by PSS, in the narrow sense replication, the earnings equation model proposed by Chan, Savage, and Whittaker (1995) is used. However, for the wide sense replication, this study uses the variable specification suggested by OBR because it provides a more recent approach concerning the data. For both cases, this study uses the real wage equation used in PSS's study, which is derived from Darby and Wren-Lewis (1993). This treatment enables a direct comparison between the two replications.

The analysis is based on a univariate framework and uses an ARDL model specification, as shown in Pesaran and Shin (1999). For an $ARDL(p, q_1, ..., q_k)$ model with k independent variables, the general form is

$$y_t = c_0 + c_1 t + \sum_{i=1}^{p} b_{y,i} y_{t-i} + \sum_{j=1}^{k} \sum_{l=0}^{q_j} b_{j,l} x_{j,t-l} + \epsilon_t \quad (1)$$

The formula for the unrestricted ECM that derives from the above-stated ARDL model is as follows:

$$\Delta y_t = c_0 + c_1 t + \pi_y y_{t-1} + \sum_{j=1}^{k} \pi_j x_{j,t-1} + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i} + \sum_{j=1}^{k} \sum_{l=1}^{q_j - 1} \psi_{j,l} \Delta x_{j,t-l} + \sum_{j=1}^{k} \omega_j \Delta x_{j,t} + \epsilon_t \quad (2)$$

The restricted ECM has the following specification:

$$\Delta y_t = c_0 + c_1 t + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i} + \sum_{j=1}^{k} \sum_{l=1}^{q_j - 1} \psi_{j,l} \Delta x_{j,t-l} + \sum_{j=1}^{k} \omega_j \Delta x_{j,t} + \pi_y ECT_t + \epsilon_t \quad (3)$$

In Equations (1), (2), and (3), c_0 and c_1 are the parameters of the deterministic terms of the constant and linear trends, respectively. In PSS, five model specifications are presented—Cases I–V.

With respect to Equation (3), ECT_t is the cointegrating equation (lagged once). For each of the five cases mentioned above, the following restrictions apply:

$$
\begin{aligned}
&\textbf{Case I:}\\
&c_0 = c_1 = 0\\
&ECT_t = y_{t-1} - (\sum_{j=1}^{k} \theta_j x_{j,t-1})\\
&\textbf{Case II:}\\
&c_0 = c_1 = 0\\
&ECT_t = y_{t-1} - (\mu + \sum_{j=1}^{k} \theta_j x_{j,t-1})\\
&\textbf{Case III:}\\
&c_1 = 0\\
&ECT_t = y_{t-1} - (\sum_{j=1}^{k} \theta_j x_{j,t-1})\\
&\textbf{Case IV:}\\
&c_1 = 0\\
&ECT_t = y_{t-1} - (\delta(t-1) + \sum_{j=1}^{k} \theta_j x_{j,t-1})\\
&\textbf{Case V:}\\
&ECT_t = y_{t-1} - (\sum_{j=1}^{k} \theta_j x_{j,t-1})
\end{aligned}
\tag{4}
$$

The bounds test is a Wald test on the estimated parameters of the model, as described in PSS. It is also important to note that the information criteria used; Akaike information criterion (AIC) and Schwarz Bayesian criterion (SBC), are those that are described in the notes of Table I in PSS, and they are calculated in such a way that they must be maximized.

E. Narrow Sense Replication

E.1. Model Specification and Test for Cointegration

Following the same steps, Table I in PSS's study is reproduced and presented in Table 4. This is the first part of the model selection process wherein the sufficient lag order for the dependent variable is determined based on the AIC and SBC, while considering the serial correlation tests. AIC suggests a lag order of 6 in both cases, with or without deterministic trends, while SBC suggests 1 and 4, respectively. Considering the importance of autocorrelation, X^2_{SC} suggests an order greater than 4.

The models tested herein are described in detail in PSS, as they are under certain restrictions and have some specific characteristics. First, in all the models, the variable *Prod* is restricted so that its lagged changes will not be involved in the model; however, the plain first difference will. Second, for model comparison reasons, the same sample period (1972:Q1–1997:Q4) is used to estimate all the models.

The two previous rules apply for all models for the rest of the narrow replication. Finally, for the analysis in Table 4, the ARDL(p) order specification described in PSS is used. This is simply an $ARDL(p, p, ..., p)$, whereby every order is the same (p) for every variable[9]. The results in Table 4 are based on the conditional unrestricted ECM.

It must be highlighted that this study replicates 55 out of 56 results of the table, except for an inconsistency in the results of $X^2_{SC}(4)$ for $p = 7$. Nonetheless, the results are qualitatively the same.

Table 4: Statistics for selecting the lag order of the earnings equation (Table I in PSS)

	With deterministic trends				Without deterministic trends			
p	AIC	SBC	$X^2_{SC}(1)$	$X^2_{SC}(4)$	AIC	SBC	$X^2_{SC}(1)$	$X^2_{SC}(4)$
1	319.33	302.14	16.86*	35.89*	317.51	301.64	18.38*	34.88*
2	324.25	301.77	2.16	19.71*	323.77	302.62	1.98	21.51*
3	321.51	293.74	0.52	17.07*	320.87	294.43	1.56	19.35*
4	334.37	301.31	3.48***	7.79***	335.37	303.63	3.41***	7.13
5	335.84	297.50	0.03	2.5	336.49	299.47	0.02	2.14
6	337.06	293.42	0.85	3.58	337.03	294.72	0.99	3.99
7	336.96	288.04	0.17	2.2	336.85	289.25	0.09	2.51

Note. Significance at 0.01, 0.05 and 0.10 is denoted by symbols $*$, $**$ and $***$, respectively.

Based on Table 4, we decided to review only the conditional unrestricted ECM of the models $ARDL(4, 1, 4, 4, 4)$, $ARDL(5, 1, 5, 5, 5)$, and $ARDL(6, 1, 6, 6, 6)$ in the next step.

The second step of the model specification selection process concerns the inclusion of the linear trend, based on whether the level relationship (cointegration) hypothesis is supported, using the bounds F- and t-test proposed in PSS.

The results from Table II in the PSS study are identical to those presented in Table 5. This suggests that there may be a level relationship using the conditional unrestricted ECM of the underlying $ARDL(5, 1, 5, 5, 5)$ for the specification without the deterministic linear trend, as it passes both the F- and the t-test.

The results in Table 5 are based on the asymptotic critical value bounds. Moreover, the exact sample critical value bounds in PSS, referring to a sample of T=104, k=4, and a 5% level of significance for the F_{IV} and F_V statistics, have been successfully replicated. This study estimates the exact sample critical value bounds for Case IV, being (3.19, 4.16), identical to those in PSS. Furthermore, the estimates for Case V of the exact sample critical value bounds are (3.63, 4.78), which only differ from PSS's values by 0.02[10].

E.2. Parsimonious Model and Long-Run Relationship

The next step includes the estimation of the level relationship and level effects (long-run

[9] Note that the restriction on $Prod$ turns the order into $ARDL(p, 1, ..., p)$.

[10] In the applied simulation, the precision is raised using 70,000 iterations. A small difference in the results is expected in such simulations owing to the different random number generation, iterations, etc.

Table 5: F- and t-statistics for testing the existence of a levels earnings equation (Table II in PSS)

	With deterministic trends			Without deterministic trends	
p	F_{IV}	F_V	t_V	F_{III}	t_{III}
4	2.99^a	2.34^a	-2.26^a	3.63^b	-3.02^b
5	4.42^c	3.96^b	-2.83^a	5.23^c	-4.00^c
6	4.78^c	3.59^b	-2.44^a	5.42^c	-3.48^b

Note. [a] indicates that the statistic lies below the lower 0.05 bound. [b] indicates that the statistic lies between the lower and the upper 0.05 bound. [c] indicates that the statistic lies above the upper 0.05 bound.

multipliers). For this, a more parsimonious specification is used. The restrictions mentioned before for the model structure are maintained (now only estimating the model with a deterministic constant but not a linear trend), except that the orders of the underlying ARDL model are allowed to vary freely from 0 to 7.

All possible specifications in the given range have been tested, which is a set of $7^5 = 16807$ models[11], thus confirming that, as in the study by PSS, the more suitable specification is $ARDL(6, 1, 5, 4, 5)$[12], according to AIC.

Equation 5 corresponds to the level relationship in Equation (31) in PSS, where all the estimations are perfectly matched. The delta method is used to estimate the standard errors of the coefficients (long-run multipliers) (Pesaran and Shin 1999).

$$\begin{aligned} w_t =&1.063\ Prod_t - 0.105\ UR_t - 0.943\ Wedge_t + 1.481\ Union_t + 2.701 \\ &(0.050) \qquad (0.034) \qquad (0.265) \qquad (0.311) \qquad (0.241) \\ &[0.000] \qquad [0.003] \qquad [0.001] \qquad [0.000] \qquad [0.000] \end{aligned} \tag{5}$$

E.3. Error Correction Model and Diagnostics (Narrow)

Finally, both the unrestricted and restricted ECMs have been successfully replicated, conditional on the underlying $ARDL(6, 1, 5, 4, 5)$ model, as presented in Tables 6 and 7, respectively. The results presented in Table III of PSS, appear in Table 6 along with the lagged variables in levels.

Additionally, the regression information, such as $\bar{R}^2$, $\hat{\sigma}$, AIC, SBC, and most of the post-estimation tests, are successfully reproduced. Two differences are found for Ramsey's RESET test for functional form and the Breusch–Pagan test for homoscedasticity and can be attributed to the various versions of the tests applied. Nevertheless, the results are qualitatively the same.

[11]This can be narrowed down to $7^3 * 6 = 2058$, considering that p spans between 1 and 6, and that q spans between 0 and 6 in general, except for q_{Prod}, which is fixed at 1.

[12]Note that in PSS, there is a typo across the paper that refers to this model specification as $ARDL(6, 0, 5, 4, 5)$, even though all their calculations are based on the $ARDL(6, 1, 5, 4, 5)$ model.

Table 6: Unrestricted equilibrium correction form of the ARDL(6,1,5,4,5) earnings equation (comparable to Table III in PSS)

Regressor	Coefficient	Standard error	p-value
w_{t-1}	-0.229	0.0586	0.000
Δw_{t-1}	-0.417	0.0974	0.000
Δw_{t-2}	-0.328	0.1089	0.004
Δw_{t-3}	-0.523	0.1043	0.000
Δw_{t-4}	-0.133	0.0892	0.140
Δw_{t-5}	-0.197	0.0807	0.017
$\Delta Prod_t$	0.315	0.0954	0.001
ΔUR_t	0.003	0.0083	0.683
ΔUR_{t-1}	0.016	0.0119	0.196
ΔUR_{t-2}	0.003	0.0118	0.797
ΔUR_{t-3}	0.028	0.0113	0.014
ΔUR_{t-4}	0.027	0.0122	0.031
$\Delta Wedge_t$	-0.297	0.0534	0.000
$\Delta Wedge_{t-1}$	-0.048	0.0592	0.417
$\Delta Wedge_{t-2}$	-0.093	0.0569	0.105
$\Delta Wedge_{t-3}$	-0.188	0.0560	0.001
$\Delta Union_t$	-0.969	0.8169	0.239
$\Delta Union_{t-1}$	-2.915	0.8395	0.001
$\Delta Union_{t-2}$	-0.021	0.9023	0.981
$\Delta Union_{t-3}$	-0.101	0.7805	0.897
$\Delta Union_{t-4}$	-1.995	0.7135	0.007
Intercept	0.619	0.1554	0.000
$D7475_t$	0.029	0.0063	0.000
$D7579_t$	0.017	0.0063	0.009
$Prod_{t-1}$	0.244	0.0662	0.000
UR_{t-1}	-0.024	0.0076	0.002
$Wedge_{t-1}$	-0.216	0.0582	0.000
$Union_{t-1}$	0.339	0.1020	0.001

$\bar{R}^2$ = 0.5589, $\hat{\sigma}$ =0.0083, AIC = 339.57, SBC = 302.55,
$\chi^2_{SC}(4) = 8.74[0.068]$, $\chi^2_{FF}(1) = 3.68[0.059]$,
$\chi^2_N(2) = 0.01[0.993]$, $\chi^2_H(1) = 0.33[0.566]$

Note. Regression statistics as described under Table III in PSS.

Table 7: Restricted equilibrium correction form of the ARDL(6,1,5,4,5) earnings equation

Regressor	Coefficient	Standard error	p-value
$\hat{u}_{t-1}$	-0.229	0.0424	0.000
Δw_{t-1}	-0.417	0.0909	0.000
Δw_{t-2}	-0.328	0.0967	0.001
Δw_{t-3}	-0.523	0.0955	0.000
Δw_{t-4}	-0.133	0.0820	0.109
Δw_{t-5}	-0.197	0.0755	0.011
$\Delta Prod_t$	0.315	0.0876	0.001
ΔUR_t	0.003	0.0070	0.629
ΔUR_{t-1}	0.016	0.0098	0.116
ΔUR_{t-2}	0.003	0.0102	0.767
ΔUR_{t-3}	0.028	0.0100	0.006
ΔUR_{t-4}	0.027	0.0101	0.010
$\Delta Wedge_t$	-0.297	0.0502	0.000
$\Delta Wedge_{t-1}$	-0.048	0.0517	0.353
$\Delta Wedge_{t-2}$	-0.093	0.0534	0.084
$\Delta Wedge_{t-3}$	-0.188	0.0537	0.001
$\Delta Union_t$	-0.969	0.7140	0.179
$\Delta Union_{t-1}$	-2.915	0.8087	0.001
$\Delta Union_{t-2}$	-0.021	0.8604	0.980
$\Delta Union_{t-3}$	-0.101	0.7481	0.893
$\Delta Union_{t-4}$	-1.995	0.6547	0.003
Intercept	0.619	0.1124	0.000
$D7475_t$	0.029	0.0051	0.000
$D7579_t$	0.017	0.0046	0.000

$\bar{R}^2$ = 0.581, $\hat{\sigma}$ =0.0081, AIC = 343.57, SBC = 311.84,
$\chi^2_{SC}(4) = 8.64[0.071]$, $\chi^2_{FF}(1) = 2.14[0.148]$,
$\chi^2_{N}(2) = 0.01[0.993]$, $\chi^2_{H}(1) = 0.33[0.566]$

Note. Regression statistics as described under Table III in PSS.

F. Wide Sense Replication

One of the advantages of the bounds test is that one does not have to know the exact order of the integration for each variable. However, this has two limitations. First, the dependent variable must be $I(1)$, and second, no variable should be $I(2)$ or greater (Pesaran *et al.* 2001). Therefore, in the first stage, the analysis evaluates whether the data violates these assumptions. Table 8 presents the results of the Augmented Dickey–Fuller (ADF) and the Phillips–Perron (PP) tests for unit root, for each variable in levels, under three different scenarios vis-a-vis the deterministic intercept and trend. Table 9 shows the same information for the first differences in the variables.

Combining the information in Tables 8 and 9, $w \sim I(1)$, and no variable is $I(2)$.

Table 8: Unit root tests on levels

	ADF test			PP test		
variable	none	intercept	int. and trend	none	intercept	int. and trend
w	2.83 (1)	-1.28 (1)	-1.67 (0)	0.12	-1.67	-7.04
	(0.999)	(0.638)	(0.762)	(0.718)	(0.805)	(0.65)
Prod	3.41 (4)	-2.47 (4)	-0.52 (4)	0.25	-2.24	-3.74
	(1)	(0.124)	(0.982)	(0.745)	(0.739)	(0.901)
UR	-0.31 (6)	-2.03 (6)	-2.28 (6)	-0.13	-4.2	-4.56
	(0.573)	(0.272)	(0.441)	(0.661)	(0.515)	(0.839)
Wedge	0.87 (5)	-0.76 (5)	-1.48 (5)	0.1	-2.91	-7.69
	(0.897)	(0.828)	(0.834)	(0.713)	(0.662)	(0.601)
Union	-1.73 (5)	-8.88 (5)	-8.23 (5)	-0.22	-6.57	-5.41
	(0.08)	(0)	(0)	(0.641)	(0.367)	(0.774)
UnionR	2.3 (4)	0.3 (4)	-2.65 (4)	0.95	0.74	-6.1
	(0.995)	(0.978)	(0.257)	(0.902)	(0.982)	(0.722)

Notes. For each deterministic type of the Augmented Dickey-Fuller tests there is the test statistic. Next to the test statistic and inside parentheses there is the number of lags included in the ADF regression. The corresponding p-value is presented in the parentheses below the test statistic.

Moreover, for each type of the Phillips-Perron test, there is the test statistic and the p-value in parentheses. The lags used in the underlying regression are 4 for every test.

Some p-values can appear as 1 or 0 as a result of the roundings.

Table 9: Unit root tests on first differences

	ADF test			PP test		
variable	none	intercept	int. and trend	none	intercept	int. and trend
Δw	-14.63 (0)	-15.18 (0)	-15.2 (0)	-219.59	-217.89	-217.84
	(0)	(0)	(0)	(0.01)	(0.01)	(0.01)
$\Delta Prod$	-8.01 (1)	-12.57 (0)	-12.79 (0)	-154.49	-167.92	-172.31
	(0)	(0)	(0)	(0.01)	(0.01)	(0.01)
ΔUR	-4.25 (5)	-4.23 (5)	-5.44 (4)	-70.62	-70.58	-75.56
	(0)	(0.001)	(0)	(0.01)	(0.01)	(0.01)
$\Delta Wedge$	-5.36 (4)	-5.43 (4)	-5.51 (4)	-199.52	-199.11	-197.87
	(0)	(0)	(0)	(0.01)	(0.01)	(0.01)
$\Delta Union$	-2.71 (4)	-2.87 (4)	-3.79 (7)	-28.86	-32.11	-40.95
	(0.007)	(0.051)	(0.019)	(0.01)	(0.01)	(0.01)
$\Delta UnionR$	-2.01 (7)	-2.69 (7)	-2.77 (7)	-195.37	-202.76	-204.39
	(0.043)	(0.077)	(0.211)	(0.01)	(0.01)	(0.01)

Note. See notes of Table 8.

F.1. Model Specification and Parsimonious Model

A similar approach to that used in PSS (narrow sense replication) is used to find a suitable model specification. First, the unrestricted ECMs conditional on the underlying $ARDL(p)$ models are estimated. Note that in contrast with the corresponding Table I in PSS (i.e., Table 4), the case without the intercept is included here. The results presented in Table 10 suggest an order of five or greater in the case of no intercept, and an order of four or greater in the other two cases.

Table 10: Statistics for selecting the lag order of the earnings equation (wide sense replication)

p	AIC	SBC	$X^2_{SC}(1)$	$X^2_{SC}(2)$	$X^2_{SC}(3)$	$X^2_{SC}(4)$	$X^2_{SC}(5)$
Without intercept or trend							
1	575.59	557.76	6.51**	6.61**	9.36**	9.53**	9.62***
2	575.14	549.21	0	0.07	4.8	5.02	5.04
3	574.97	540.93	4.66**	17.96*	17.99*	18.09*	18.18*
4	575.17	533.03	1.45	4.8***	4.87	5.87	8.35
5	575.19	524.94	0.03	1.16	1.28	2.37	2.76
6	575.18	516.83	0	2.15	2.52	2.69	3.3
7	571.86	505.41	1.28	4.54	4.96	6.08	6.17
With intercept							
1	582.32	562.86	5.98**	6.13**	10.62**	11.13**	11.16**
2	581.26	553.70	0.05	0.05	7.35***	7.78	7.78
3	582.13	546.47	2.04	15.79*	15.8*	16.19*	16.2*
4	584.45	540.69	1.48	3.43	3.59	4.37	7.53
5	582.96	531.09	1.46	2.05	2.37	2.96	3.74
6	580.51	520.54	0.28	1.03	1.55	1.78	3.16
7	576.84	508.76	1.35	3.49	4.55	4.9	5.32
With intercept and trend							
1	581.62	560.55	5.95**	6.03**	11.4*	12.36**	12.58**
2	580.49	551.31	0.14	0.18	8.67**	9.41***	9.45***
3	582.23	544.95	1.28	15.7*	15.8*	16.53*	16.53*
4	586.20	540.81	1.29	4.07	4.09	5.08	7.18
5	584.67	531.18	2.48	3.36	3.38	3.39	4.11
6	582.53	520.94	0.4	0.76	0.89	1	1.57
7	579.78	510.09	2.46	4.71***	5.01	7.42	7.47

Notes. AIC and SBC denote Akaike and Schwarz Bayesian Information Criteria not in the standard way but as defined in PSS. In addition, the columns from $\chi^2_{SC}(1)$ to $\chi^2_{SC}(5)$ refer to the test statistics of the Breusch-Godfrey LM test for serial correlation of order 1 through 5 respectively.
Significance at 0.01, 0.05 and 0.10 is denoted by symbols $*$, $**$ and $***$, respectively.

Upon reviewing the unrestricted ECMs in Table 10, it is confirmed that, in general, the lagged changes of $Prod$ are statistically insignificant across the models, as in PSS. The same behavior is observed for $Union$, except that in this case, not even the plain first difference is statistically significant. This is a strong indication that the appropriate orders for $Prod$ and $Union$ may be 1 and 0, respectively. Calculating Table 10, while considering these restrictions, leads to the exact same conclusions; thus, this table sufficiently illustrates the general picture.

The inconstant data in Figures 5, 6, 7, 8, 9, and 10 suggest that a model without any deterministic terms has a poor fit and is found to be a poor candidate. Consequently, it is not considered in the search for the best parsimonious model, thus reaching the same conclusion as in many other studies (Pesaran *et al.* 2001; Tyrvainen 1995). A global search across the grid of all the $7^4 * 6 = 14406$ candidate models (with an intercept) shows that the best model (according to AIC) is the $ARDL(4, 1, 0, 5, 0)$ model, and the second-best is $ARDL(4, 1, 3, 5, 0)$. The global search across all possible models containing an intercept and a linear trend shows that the best model is $ARDL(4, 1, 6, 5, 2)$ and the second-best is $ARDL(4, 1, 3, 5, 2)$.

F.2. Cointegration Test and Long-Run Relationship

Table 11 presents the results of the asymptotic bounds F-test and the bounds t-test (where applicable) for the best two candidate models with intercept (under Cases II and III) and the best two candidate models with intercept and trend (under Cases IV and V).

In addition, Table 12 presents the exact sample critical value bounds (for a significance level of 0.05). Such a treatment enables us to adjust the results for the real sample size of T=189 observations instead of the T=1000 (asymptotic) used in PSS. Furthermore, the exact sample p-values are calculated to determine the level of significance wherein the null hypothesis of the upper bound can be rejected. Thus, one does not have to repeat the test for each level (e.g., 1%, 5%, 10%, etc.) and can easily conclude based on the p-values. Note that the p-value concerns the upper bound; thus, rejecting the null hypothesis concludes in favor of the alternative for the existence of a possible cointegration, but it does not inform us whether the test statistics are between the two bounds or under the lower bound.

The findings presented in Tables 11 and 12 are the same, as the sample is not very small to be greatly misled by the asymptotic critical value bounds. Interpreting the results, one can see that for Case V and for both models, the null hypothesis of no-cointegration cannot be rejected by the bounds t-test. The same occurs in Case III for the $ARDL(4, 1, 3, 5, 0)$ model. Thus, it must be mentioned that the existence of a cointegrating relationship in Cases II and IV cannot be tested further using the bounds t-test.

Table 11: Asymptotic F- and t-statistics for testing the existence of a levels earnings equation (wide sense replication)

	With intercept			With intercept and trend		
ARDL(p, q_1, q_2, q_3, q_4)	F_{II}	F_{III}	t_{III}	F_{IV}	F_V	t_V
ARDL(4,1,0,5,0)	4.13^c	4.32^c	-4.03^c	-	-	-
ARDL(4,1,3,5,0)	4.01^c	4.29^c	-3.67^b	-	-	-
ARDL(4,1,6,5,2)	-	-	-	4.15^c	4.72^c	-4.07^b
ARDL(4,1,3,5,2)	-	-	-	4.28^c	5.05^c	-4.1^b

Notes. ARDL(p, q_1, q_2, q_3, q_4) stands for ARDL($p_w, q_{Prod}, q_{UR}, q_{Wedge}, q_{Union}$).
F_{II}, F_{III}, F_{IV} and F_V are the F-statistics of the bounds F-test for each case. t_{III} and t_V are the t-statistics of the bounds t-test for each case.
^{a} indicates that the statistic lies below the lower 0.05 bound. ^{b} indicates that the statistic lies between the lower and the upper 0.05 bound. ^{c} indicates that the statistic lies above the upper 0.05 bound.
The asymptotic critical value bounds are those from PSS, assuming T=1000 observations.

Table 12: Exact sample F and t critical value bounds and p-values for testing the existence of a levels earnings equation (wide sense replication)

	With intercept		
ARDL(p, q_1, q_2, q_3, q_4)	F_{II}	F_{III}	t_{III}
ARDL(4,1,0,5,0)	[2.62, 3.56] 0.0193	[2.92, 4.09] 0.0360	[−2.86, −3.99] 0.0462
ARDL(4,1,3,5,0)	[2.62, 3.56] 0.0237	[2.92, 4.09] 0.0370	[−2.86, −3.99] 0.0988
	With intercept and trend		
ARDL(p, q_1, q_2, q_3, q_4)	F_{IV}	F_V	t_V
ARDL(4,1,6,5,2)	[3.13, 4.05] 0.0432	[3.56, 4.68] 0.0473	[−3.41, −4.36] 0.0918
ARDL(4,1,3,5,2)	[3.13, 4.05] 0.0347	[3.56, 4.68] 0.0296	[−3.41, −4.36] 0.0862

Notes. ARDL(p, q_1, q_2, q_3, q_4) stands for ARDL($p_w, q_{Prod}, q_{UR}, q_{Wedge}, q_{Union}$).
Under F_{II}, F_{III}, F_{IV} and F_V are the lower and upper exact sample critical value bounds for the F-test for each case, while under the t_{III} and t_V are the equivalents for the t-test. These critical value bounds correspond to the 0.05 level of significance.
Under the critical value bounds there are the exact sample p-values, referring to the null hypothesis for the upper bound.
The F- and t-statistics are the same that appear in the Table 11. For each stochastic simulation 70000 iterations were used and the exact sample used was T=189 observations.

Figures 13 and 14 present the real wages (w) compared with the long-run relationships of these four candidate models for each case. Note that under Cases III–V, where the intercept is not restricted (and thus not included in the levels relationship), the level relationships diverge from the real wages.

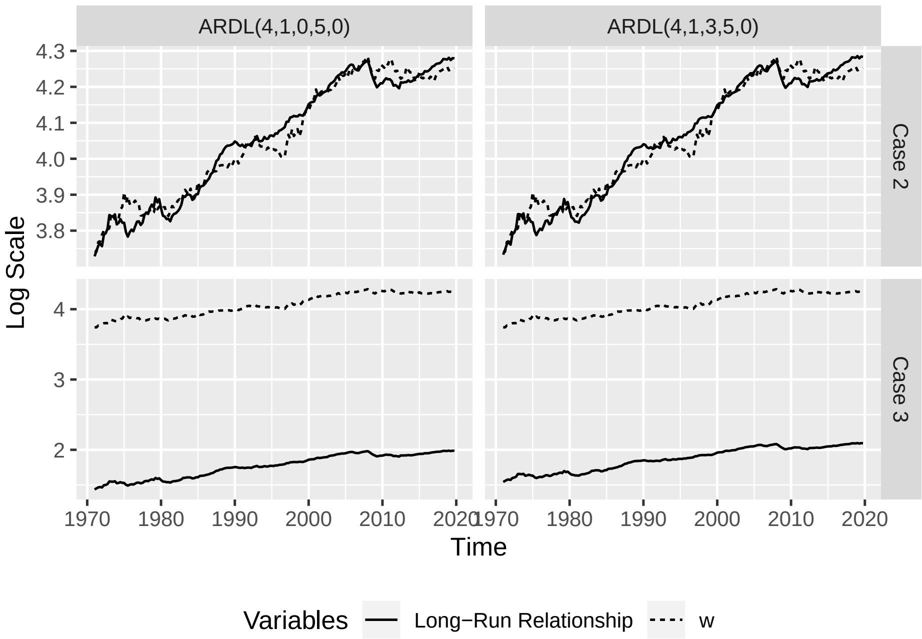

Figure 13: Real wages (w) and long-run relationship (cointegrating equation) for models with intercept

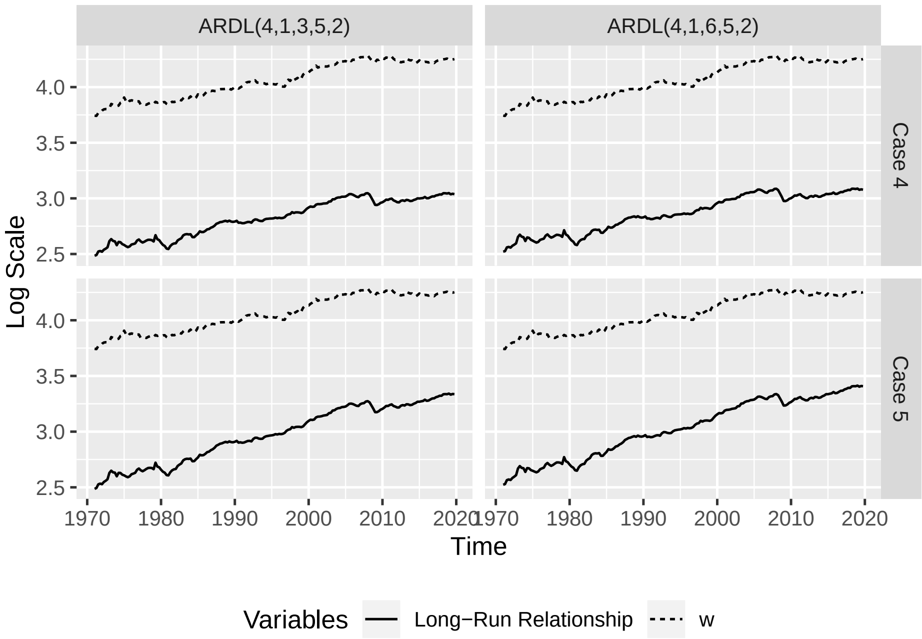

Figure 14: Real wages (w) and long-run relationship (cointegrating equation) for models with intercept and trend

Consequently, based on the AIC and the results from Tables 10 and 11, the remaining analysis is based on the conditional ECM of the $ARDL(4,1,0,5,0)$ model in Case II[13].

Equation 6 presents the level relationship based on the selected model. Compared with Equation 5, few differences are reported.

$$\begin{aligned} w_t = & 0.855\ Prod_t - 0.110\ UR_t + 0.174\ Wedge_t + 0.017\ Union_t + 2.292 \\ & (0.072) \quad (0.032) \quad (0.153) \quad (0.418) \quad (0.470) \\ & [0.000] \quad [0.001] \quad [0.258] \quad [0.968] \quad [0.000] \end{aligned} \tag{6}$$

First, the variable $Wedge$ is statistically insignificant. This is not a surprising result, as there are mixed evidence and arguments in the literature about whether there is a long-run wedge effect on real wages (Pesaran *et al.* 2001). In the study of Office for Budget Responsibility (OBR) (2013) for the period 1972:Q4–2007:Q4, the same results of no long-run wedge effect are found. In the study of Tyrvainen (1995) for the period 1969–1991, although a multivariate approach is followed and the wedge effect is not estimated directly, but through its various tax rate components, the wedge effect is estimated to be approximately 0.25; this is a relatively low wage resistance as it is closer to zero than to unity. In the wide replication case, with the coefficient being statistically equal to zero, it can be interpreted that the firms set the wages, while tax payment fall almost entirely on the employees.

$Union$ is also insignificant, probably because over 80% of the variable's span has almost no variability. OBR uses the same measure for the variable $Union$ and estimates a greater coefficient (0.141) but not as large as the one estimated by PSS (-0.943).

Though it has a correct sign, $Prod$ has a smaller coefficient compared with that in PSS[14]. In many studies, it is set a priori to unity (Office for Budget Responsibility (OBR) 2013; Tyrvainen 1995). Herein, testing whether the coefficient of $Prod$ is statistically equal to unity is marginally rejected (using $a = 0.05$) with $p - value = 0.046$[15]. Other studies have estimated it as high as 1.26 (Arestis and biefang-frisancho Mariscal 1994) for the period 1966:Q1–1989:Q2, or as low as 0.75 (Hall 1986) for the period 1965:Q3–1984:Q3.

The long-run coefficient of UR has a correct sign and is very close to that in the narrow sense replication. In fact, the majority of the studies estimate the same magnitude (and sign) of unemployment elasticity; for example, Tyrvainen (1995) and Arestis and biefang-frisancho Mariscal (1994) calculate this as -0.10 and -0.11, respectively. A notable deviation from this is Office for Budget Responsibility (OBR) (2013), which estimates this as low as -0.02. Other studies, such as Hall (1986), do not use a log–log regression, and the estimated coefficient refers to semi-elasticity.

F.3. Error Correction Model and Diagnostics (Wide)

Table 13 includes the results of the conditional unrestricted ECM regression of the underlying $ARDL(4,1,0,5,0)$ model, some diagnostic tests (all tests pass), and information about the regression. Note that the coefficient of the lagged real wages (w) is the error correction

[13] At this point, there is nothing wrong with continuing the analysis with an equivalently suitable model specification such as $ARDL(4,1,3,5,0)$, which has some interesting properties, as discussed in Subsection G.3

[14] See Subsection G.1 for more details about $Wedge$ and $Prod$ and subsection F.4 for the discussion about $Union$

[15] The corresponding 95% confidence interval is [0.711, 0.999].

coefficient. In wide sense replication, this is estimated as -0.125, which is approximately half of the estimate in narrow replication, meaning that it would take eight periods (quarters) or equivalently two years to return to equilibrium.

Table 13: Unrestricted equilibrium correction form of the ARDL(4,1,0,5,0) earnings equation (wide sense replication)

Regressor	Coefficient	Standard error	p-value
Intercept	0.286	0.0744	0.000
w_{t-1}	-0.125	0.0309	0.000
$Prod_{t-1}$	0.107	0.0289	0.000
UR	-0.014	0.0039	0.000
$Wedge_{t-1}$	0.022	0.0179	0.229
$Union_{t-1}$	0.002	0.0520	0.968
Δw_{t-1}	-0.102	0.0657	0.121
Δw_{t-2}	0.023	0.0651	0.723
Δw_{t-3}	-0.143	0.0657	0.031
$\Delta Prod_t$	0.577	0.0935	0.000
$\Delta Wedge_t$	-0.256	0.0491	0.000
$\Delta Wedge_{t-1}$	-0.059	0.0551	0.289
$\Delta Wedge_{t-2}$	0.036	0.0539	0.508
$\Delta Wedge_{t-3}$	-0.123	0.0528	0.021
$\Delta Wedge_{t-4}$	0.137	0.0480	0.005
$D7475_t$	0.013	0.0051	0.012
$D7579_t$	0.000	0.0041	0.913

$\bar{R}^2 = 0.3287$, $\hat{\sigma}$ =0.0102, AIC = 590.98, SBC = 563.42,
$\chi^2_{SC}(4) = 2.61[0.624]$, $\chi^2_{FF}(1) = 0.08[0.773]$,
$\chi^2_N(2) = 1.67[0.434]$, $\chi^2_H(1) = 2.04[0.153]$

Notes. Results of the conditional unrestricted ECM regression. $\bar{R}^2$ is the adjusted coefficient of determination, $\hat{\sigma}$ is the regression's standard error and AIC and SBC are the Akaike and Schwarz Bayesian Information Criteria as defined in PSS. $\chi^2_{SC}(4)$ is the test statistic of the Breusch-Godfrey LM test for serial correlation of order up to 4 (tests for lower and greater orders pass too). $\chi^2_{FF}(1)$ is the test statistic of the RESET test for functional form misspecification, based on the power of 2 (tests for higher single and multiple powers pass too). $\chi^2_N(2)$ is the Jarque-Bera test statistic for normality and $\chi^2_H(1)$ is the test statistic of the Breusch-Pagan test for heteroskedasticity. p-values given in [•].

It is also important to test the stability of the conditional ECM coefficients. Figure 15 clearly shows that the model also passes these tests.

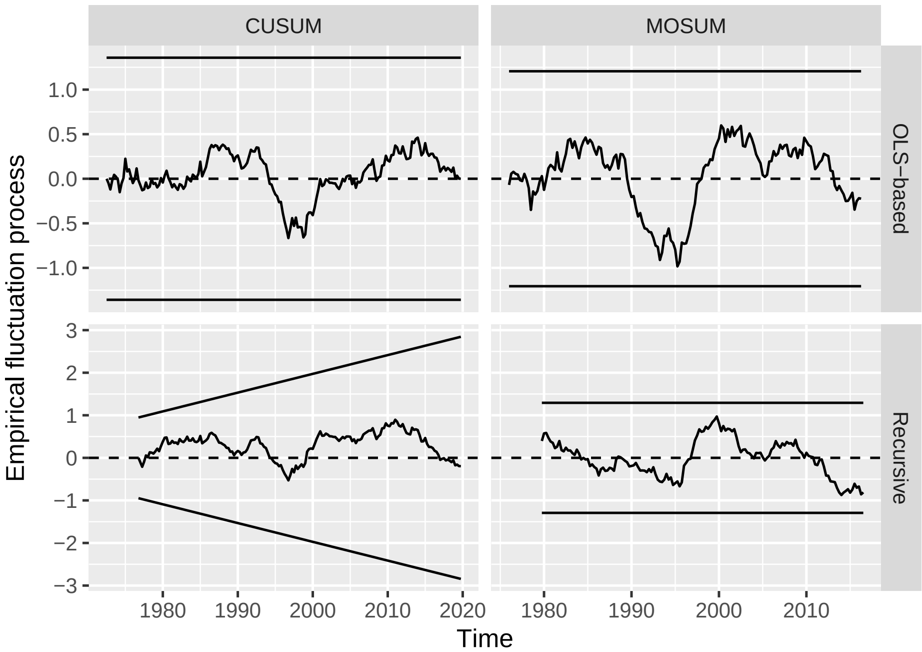

Figure 15: Stability tests for the stability of the conditional ECM's coefficients ($alpha = 0.05$)

F.4. Extended Wide Replication Model

The dummies used in PSS for the periods 1974:Q1–1975:Q4 and 1975:Q1–1979:Q4 can be interpreted as a way to capture the vast increase in trade union membership that led to strike actions from 1974, including the Grunwick dispute in 1976-1978 and the Winter of Discontent in 1978-1979. This can be seen in the corresponding years in Figure 3. In fact, the same $Union$ measure is used by OBR, noting that "it is used as a proxy for structural changes in the labor market".

The possibility of introducing extra dummy variables to capture the potential effects of policy changes or events, such as the global financial crisis of 2007 in the short-run, is also considered; the findings show no statistically or practically significant effects. This seems logical, as there have been no major events affecting union power or other instances since then. Indeed, the impact of the financial crisis is clear (i.e., the flattened trend observed in Figures 5 and 6 after 2007), but the dummy variables in the ARDL framework enter the short-run relationship and capture potential steep changes in the differenced variables, but not in the level equation.

For wide sense replication, we use the same measure for $Union$ as in PSS and OBR as a proxy for labor market changes.

Herein, the wide sense replication expands even further, using another measure for $Union$ (for distinction, it is called $UnionR$) that presents a gradual decline over time (see Figure 10), as it measures the union membership rate, rather than a flat line representing no major events. This is probably a more suitable approach, as it more accurately describes the continuous decline of union membership over the last 35 years. Another difference in the extended model is that the dummy $D7579$ is dropped, as it has no significant effect[16].

Henceforth, the wide sense replication modeling presented before will be referred to as the "wide sense replication", and the one presented in this subsection will be denoted as the "extended wide sense replication". The major findings are discussed hereby, following the same modeling strategy as in the case of the previous wide replication.

First, upon calculating the same statistics as in Table 10 for the new model specification, no practical differences are noticeable, resulting in the same interpretation. Interestingly, a global search across all possible models reveals an even more stable model selection than the previous wide sense replication model [17]. It was viable to choose Orders 4, 1, 0, and 5 for variables w, $Prod$, UR, and $Wedge$, respectively, but it was indifferent to choose between Orders 3–6 for the variable $UnionR$. Therefore, we decided to continue the analysis using the $ARDL(4, 1, 0, 5, 5)$ model based on the AIC.

The results from the bound test, accounted for Cases II–V, supported those for wide replication (Table 11), and even made a stronger suggestion in favor of the existence of a level relationship, as every test was able to reject the null hypothesis. As in the previous model, the case under focus here is also Case II, wherefore the long-run relationship fit is similar to that of the selected model in Figure 13.

Having selected an adequate model ($ARDL(4, 1, 0, 5, 5)$), the long-run multipliers presented in Equation 7 are estimated. The long-run results in this extended wide sense replication case are like those in the previous model. The elasticity of productivity is closer to unity than before, while that of unemployment is slightly lower in magnitude. Larger changes are

[16] Even if it is included in the model as in wide replication, it has an infinitesimal effect on the results
[17] See Subsection G.2

found in the elasticities of $Wedge$ and $Union$, -39% and +147%, respectively, but they are still statistically insignificant.

Notable benefits of using $UnionR$ are the dynamics in the volatility that, as opposed to the flat line of variable $Union$, contribute to the adjustment in the long-run with an estimated $\pi_y = -0.175$.

$$\begin{aligned} w_t =& 0.952\ Prod_t - 0.092\ UR_t + 0.107\ Wedge_t + 0.042\ UnionR_t + 1.987 \\ & (0.124) \quad (0.029) \quad (0.110) \quad (0.074) \quad (0.259) \\ & [0.000] \quad [0.002] \quad [0.333] \quad [0.572] \quad [0.000] \end{aligned} \tag{7}$$

G. Robustness Analysis

In this section, a deeper dive is presented, analyzing the earnings equation model of the UK economy, compared to the original paper, as PSS focus more on methodology. Through a series of robustness analyses and modeling using various setups, additional insights into the dynamics of the UK earnings equation of wide sense replication are gained.

G.1. Long-Run Effects Differences Between the Two Replications

One may wonder whether the different long-run effects are due to the extended period or data dissimilarities[18]. This is a crucial question that needs to be answered. To address this, two scenarios are set up, running the same procedure.

(a) First, using the wide sense data for the period 1971:Q1–1997:Q4[19].

(b) Thereafter, the same is done, but the deflator is changed. PYNONG (the deflator used in narrow sense replication) is used instead of PGVA. The remaining data are those used in wide sense replication, but for the period 1971:Q1–1997:Q4. The deflator is part of the dependent variable (real wages) and the independent variable $Wedge$.

The variable $Wedge$ under Scenario (b) is very similar to the same variable used in narrow sense replication, whereas the version that is used in the wide sense replication has different scales and trends, albeit similar variations, as shown in Figure 16.

Regarding the dependent variable (w), under Scenario (b), it is scaled closer to the real wages in the narrow sense replication, but its variability is similar to the real wages used in the wide sense replication, as Figure 17 shows.

[18] explained in Section C.2

[19] The wide sense data do not cover the period prior to this date. As pertains to the rest of the analysis, every calculation is performed starting from 1972:Q4. The same applies to Scenario (b)

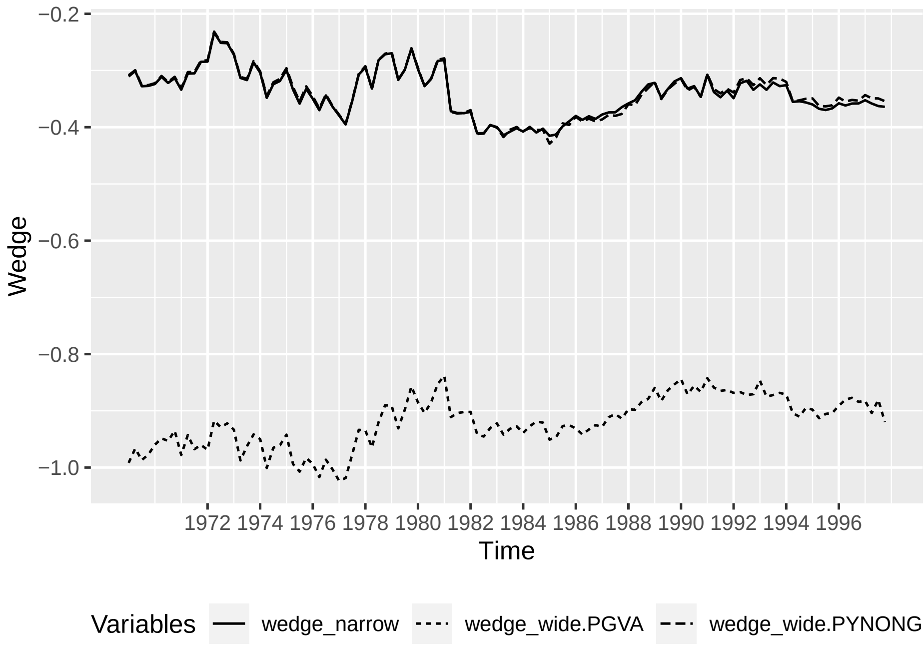

Figure 16: Wedge in narrow sense replication, scenario (a) and scenario (b)

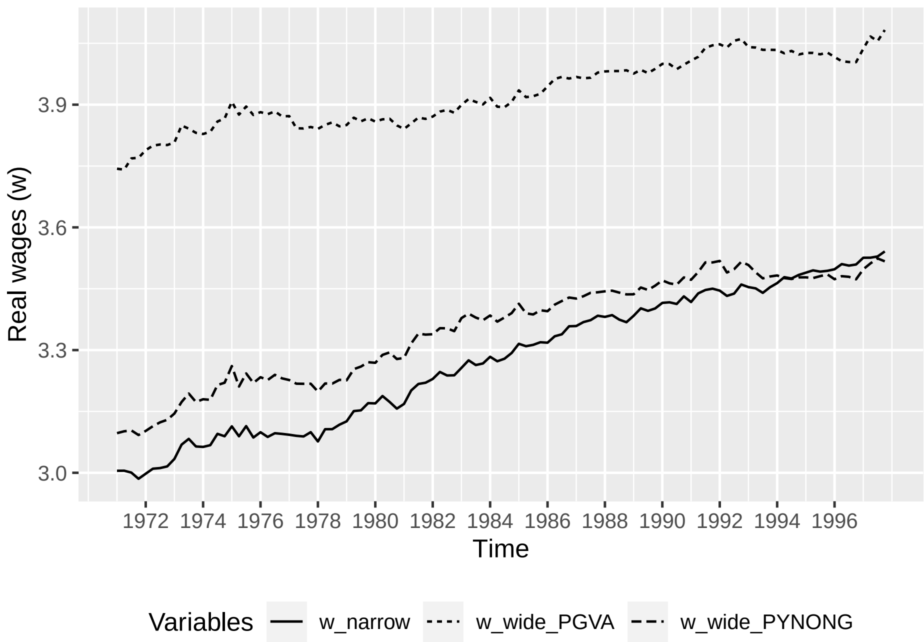

Figure 17: Real wages in narrow sense replication, scenario (a) and scenario (b)

Similarly, Tyrvainen (1995) replaces CPI with the private consumption deflator to obtain better results.

Scenario (a) concludes that the wide sense replication results are robust to changes in the period examined. The assessment is based on the following criteria:

Robustness of the model selection. The AIC-based optimization results in very similar models to those of wide sense replication (i.e., the orders of w, $Prod$, and $Union$ are 4, 1, and 0, respectively).

Conclusions regarding the cointegrating relationship. The long-run hypothesis is fully supported, as in the full-sample wide sense replication.

The long-run effects. These values are very close to those from wide sense replication and are practically equivalent.

Equations 8 (ARDL(4,1,3,4,0)), 9 (ARDL(4,1,4,4,0)), and 10 (ARDL(4,1,1,5,0)) depict these findings using three representative models.

$$\begin{aligned} w_t =\ & \underset{\substack{(0.044)\\ [0.000]}}{0.751}\ Prod_t - \underset{\substack{(0.022)\\ [0.008]}}{0.058}\ UR_t + \underset{\substack{(0.137)\\ [0.867]}}{0.023}\ Wedge_t - \underset{\substack{(0.200)\\ [0.373]}}{0.179}\ Union_t + \underset{\substack{(0.241)\\ [0.000]}}{2.155} \end{aligned} \tag{8}$$

$$\begin{aligned} w_t =\ & \underset{\substack{(0.044)\\ [0.000]}}{0.760}\ Prod_t - \underset{\substack{(0.022)\\ [0.006]}}{0.064}\ UR_t - \underset{\substack{(0.137)\\ [0.975]}}{0.004}\ Wedge_t - \underset{\substack{(0.200)\\ [0.452]}}{0.151}\ Union_t + \underset{\substack{(0.234)\\ [0.000]}}{2.139} \end{aligned} \tag{9}$$

$$\begin{aligned} w_t =\ & \underset{\substack{(0.048)\\ [0.000]}}{0.763}\ Prod_t - \underset{\substack{(0.022)\\ [0.030]}}{0.048}\ UR_t - \underset{\substack{(0.148)\\ [0.945]}}{0.010}\ Wedge_t - \underset{\substack{(0.209)\\ [0.216]}}{0.261}\ Union_t + \underset{\substack{(0.256)\\ [0.000]}}{2.019} \end{aligned} \tag{10}$$

Scenario (b) concludes that the results match those of narrow sense replication more than those of wide sense replication. The same criteria apply here.

Robustness of model selection. The empirical finding that the order of $Prod$ is 1, which is supported both in narrow and wide sense replications, is not established here.

Conclusions regarding the cointegrating relationship. A long-run relationship is not confirmed under this scenario. The narrow sense model has a similar behavior, whereby the long-run relationship is only confirmed based on the ARDL(5,5,5,5,5) model with a test statistic (t-statistic) of -4, while the asymptotic critical value is -3.99, which barely rejects the null hypothesis.

All the long-run multipliers except that of $Prod$ are very unstable, even for similar models, and they are all statistically insignificant. Equations 11 (ARDL(6,3,5,7,6)), 12 (ARDL(6,3,5,5,6)), and 13 (ARDL(6,3,6,5,6)) show three representative models for this scenario.

$$\begin{aligned} w_t =& 0.841\ Prod_t - 1.108\ UR_t - 5.386\ Wedge_t + 3.180\ Union_t + 4.052 \\ & (0.244) \quad (1.835) \quad (8.715) \quad (5.134) \quad (4.643) \\ & [0.001] \quad [0.548] \quad [0.539] \quad [0.538] \quad [0.386] \end{aligned} \tag{11}$$

$$\begin{aligned} w_t =& 0.868\ Prod_t - 2.499\ UR_t - 11.576\ Wedge_t + 6.398\ Union_t + 7.073 \\ & (0.553) \quad (9.080) \quad (41.666) \quad (22.974) \quad (21.086) \\ & [0.121] \quad [0.784] \quad [0.782] \quad [0.781] \quad [0.738] \end{aligned} \tag{12}$$

$$\begin{aligned} w_t =& 0.754\ Prod_t - 2.283\ UR_t - 10.552\ Wedge_t + 6.185\ Union_t + 7.085 \\ & (0.700) \quad (8.393) \quad (38.394) \quad (22.429) \quad (21.330) \\ & [0.285] \quad [0.786] \quad [0.784] \quad [0.784] \quad [0.741] \end{aligned} \tag{13}$$

In addition, under Scenario (b), the long-run relationship is very volatile and unstable across time, contrariwise to the stable and less volatile long-run relationship under Scenario (a). Figures 19 and 18 clearly illustrate this.

The coefficient of $Prod$ in these scenarios and in the wide replication is lower than 1, while in the narrow sense, replication is larger. As shown in Figure 17 and also explained in Subsection C.2, the coefficient of w, between the narrow and wide replications, differs because of the deflator and slightly different measurement of the real wages.

This robustness analysis concludes that using PYNONG as the deflator makes the long-run multipliers unstable, and the long-run relationship is very volatile and unstable across time. Using PGVA as the deflator, the model selection process is more consistent, the long-run relationship hypothesis is fully supported, and the estimation of the long-run multipliers is robust regardless of the time period examined (1970:Q1–1997:Q4 or 1971:Q1–2019:Q4).

Thus, the data used in wide sense replication can model the economic concept investigated much better, thereby providing stable model estimations across time.

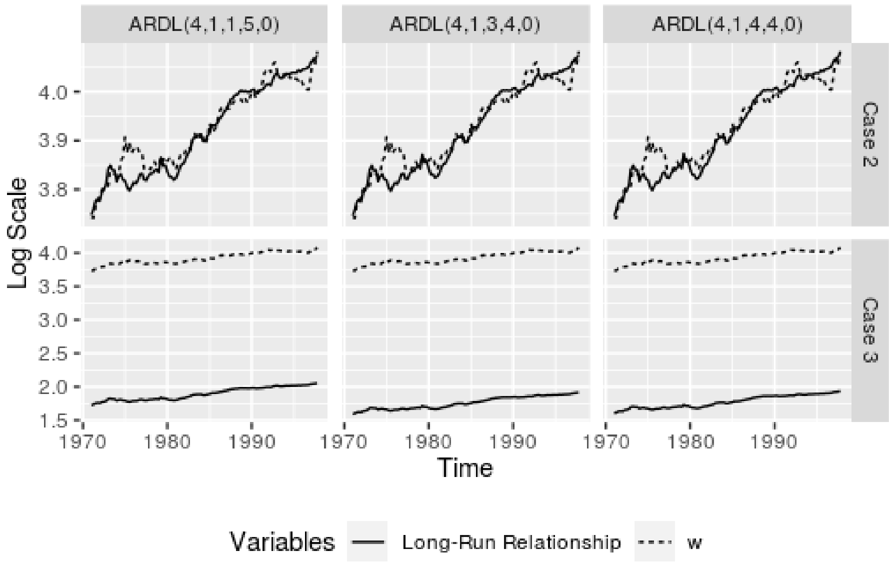

Figure 18: Long-run relationship under scenario (a)

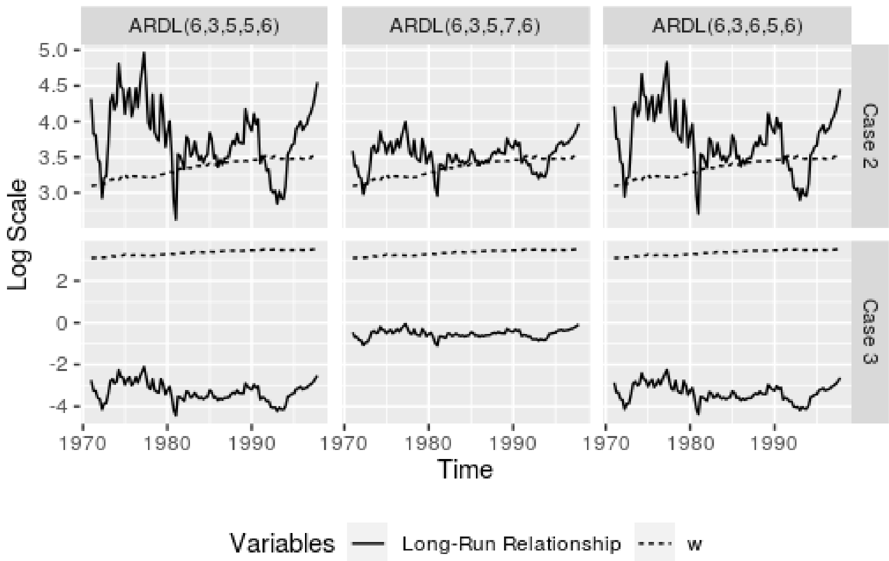

Figure 19: Long-run relationship under scenario (b)

G.2. How Sensitive are the Wide Replication Results to the Choice of Model?

The top 100 model based on the AIC (as defined in PSS, which is the same as narrow and wide sense replications use) are estimated. Thereafter, the minimum and the maximum estimated long-run coefficients and the percentage of the statistically significant variables across all models are calculated. The results are shown in Table 14, illustrating that the estimated coefficients are stable across time, as shown in Subsection G.1 and very stable across 100 different models covering a large set of specifications. In addition, all models conclusively present the same statistical significance, using α=5% for all variables.

At this point, it is evident that the variables that have a long-run effect on the earnings equation, namely $Prod$ and UR, have stable long-run effects. They also satisfy the restrictions for a well-defined earnings equation in the sense of a wage setting schedule, $0 \leq \theta_{Prod} \leq 1$ and $\theta_{UN \leq 0}$ (Tyrvainen 1995), where θ_j is the long-run coefficient, as mentioned in Equation 4.

Table 14: Summary of top 100 models

term	min	max	stat_sig_perc
Prod	0.847	0.879	100
UR	-0.125	-0.090	100
Wedge	0.052	0.208	0
Union	-0.202	0.214	0
(Intercept)	1.947	2.434	100

Note. min and max show respectively the minimum and the maximum value across the 100 models estimated and stat_sig_perc presents the percentage of each variable being statistically significant (α=5%) across all models.

G.3. Behavior of the Error Correction Coefficient and R Squared

The impact of the inclusion of the lags of ΔUR and $\Delta Union$ on $\bar{R}^2$ and the error correction coefficient π_y (see Equation (3)) is investigated in this subsection.

For $\bar{R}^2$, the results are quite straightforward. The inclusion of lags has no impact on $\bar{R}^2$ in wide sense replication. Under Scenario (b) (described in Subsection G.1), when lags for both UR and $Union$ are included (i.e., ARDL(6,3,5,7,6)), $\bar{R}^2$ becomes 0.52%, similar to PSS. Moreover, except for the difference in the dependent variable, another factor for the difference between the value of $\bar{R}^2$ in the narrow and wide replications is probably the smaller sample size. A sign of this conclusion is that including lags of UR raises the value of $\bar{R}^2$ under Scenario (a), which is in contrast with the full wide replication sample. The code with the examples that lead to this conclusion can be found in the `robustness_analysis.R` file. In Tyrvainen (1995), R^2 is equal to 55% and 76%, thus allowing for the cointegration rank to be 3 and 5, respectively; however, this explanation of the variability is expected in a multivariate analysis as feedback from the other equations is added in the equation of interest.

The value of π_y in narrow sense replication is -0.229, and in wide sense replication using the ARDL(4,1,0,5,0) model is -0.125. In the wide sense replication, its magnitude increases when the order of UR is raised (i.e., includes lags of ΔUR). For example, for the ARDL(4,1,4,5,0) model, the value of π_y is -0.163. The same also occurs when using the subsample of Scenario (a), as mentioned above. The estimated π_y for the ARDL(4,1,0,5,0) and ARDL(4,1,4,5,0) models become -0.288 and -0.391, respectively. The same also occurs under Scenario (b), but it is noteworthy that the inclusion of lags for both UR and $Union$ shrinks the estimated value of π_y to -0.074, in contrast with the effect it has on $\bar{R}^2$.

The investigation of π_y is also incorporated in the design scheme described in Subsection G.2 for the long-run coefficients. This reveals an interesting pattern for the value of π_y that depends on the order of UR (or equivalently, the lags of ΔUR). The estimated π_y from the top 100 models based on the AIC lies between -0.195 and -0.118 with a mean of -0.145. Figure 20 shows the distribution of these values. Thus, this is a bimodal distribution, and upon searching for the discriminating factor, we find that the distribution of the order of UR for the models with $\pi_y < -0.145$ is very different from that for the models with $\pi_y > -0.145$. These are shown in Figures 21 and 22, respectively. Thus, increasing the order of UR (up to a certain point) keeps the estimated π_y growing in magnitude (becoming smaller in value).

Upon using the parsimonious ARDL(4,1,0,5,0) model demonstrated in Section F, the value of π_y is -0.125. However, on estimating the ARDL(4,1,3,5,0) model, this value becomes -0.158. The choice of a model based only on certain criteria, such as AIC, is not always ideal. Therefore, adopting a less parsimonious model, such as this, which can take advantage of the additional dynamics of UR that have a significant impact on π_y, may be preferred. Tyrvainen (1995) notices a similar behavior whereby upon using long lags of UR, the value of π_y is affected because of the slow response of unemployment to changes in wages. The estimated speed of adjustment in this study is -0.7; however, as the author mentions, it is biased because of a second vector that enters the equation. For the purposes of this replication, this study adheres to the univariate framework, but a multivariate approach could inform future work to gain more insights through parallel error correction mechanisms.

Meanwhile, as mentioned in Section G.2, the long-run coefficients of the less parsimonious model (ARDL(4,1,3,5,0) in equation 14) are not practically different from those of the ARDL(4,1,0,5,0) model in Equation 6.

Interestingly, the same analysis on the extended wide replication model (using $UnionR$) does not have the same effect of lags of UR on the speed of adjustment. Moreover, the estimated $\pi_y = -0.175$ is even higher in magnitude. It is possible that the lags of UR captures some of the effects of the variation in the union membership rate, thus leading to a faster long-run adjustment.

$$\begin{aligned} w_t = & 0.858\ Prod_t - 0.108\ UR_t + 0.120\ Wedge_t - 0.041\ Union_t + 2.190 \\ & (0.059) \qquad (0.027) \qquad (0.129) \qquad (0.356) \qquad (0.418) \\ & [0.000] \qquad [0.000] \qquad [0.353] \qquad [0.908] \qquad [0.000] \end{aligned} \tag{14}$$

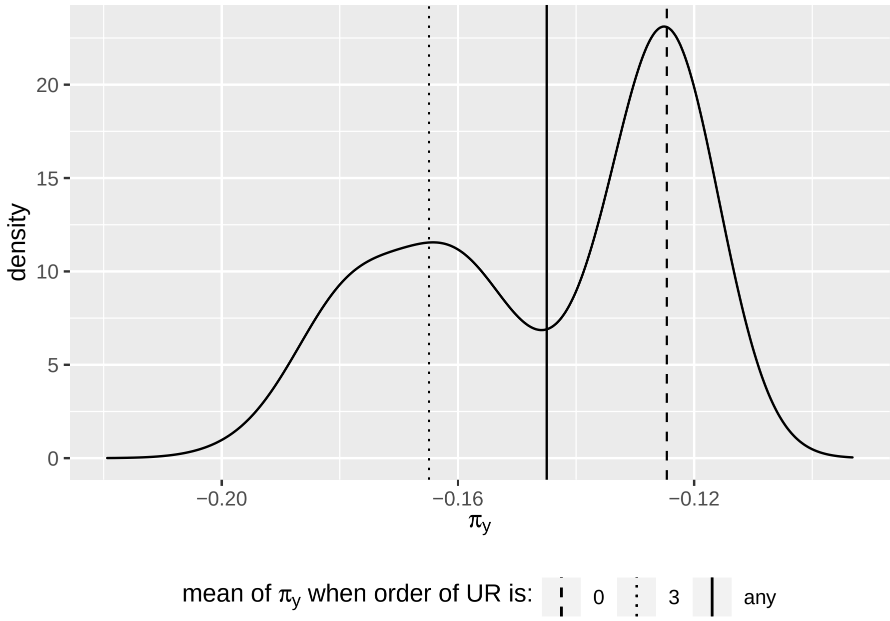

Figure 20: Density plot of π_y from the top 100 models

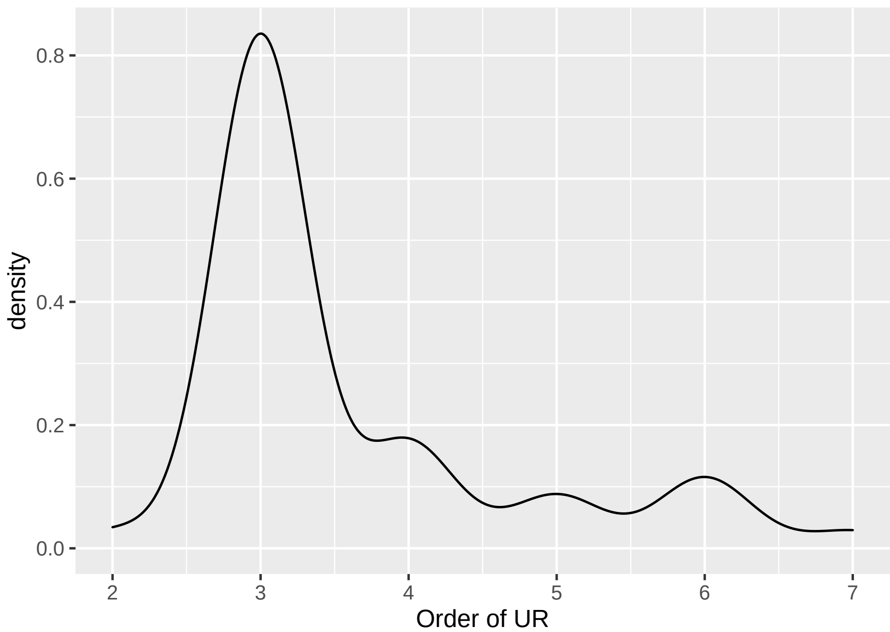

Figure 21: Distribution of orders, for models with $\pi_y < -0.145$

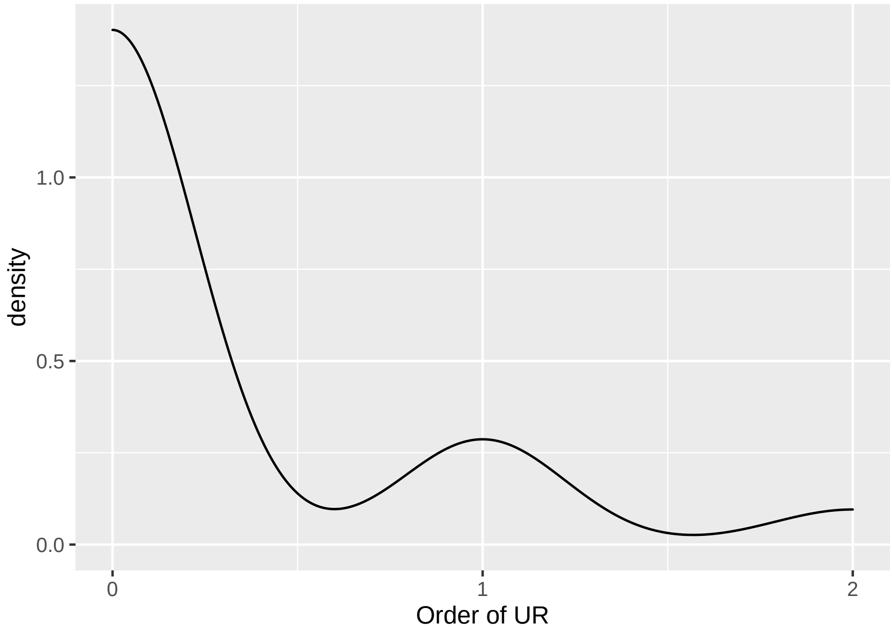

Figure 22: Distribution of orders, for models with $\pi_y > -0.145$

H. Conclusions

In the narrow sense replication, this study successfully reproduces the exact results of Pesaran *et al.* (2001) using R (R Core Team 2020) and the ARDL package (Natsiopoulos and Tzeremes 2021) to implement the proposed methodology. The wide sense replication is able to provide a fresh investigation of the UK earnings equation using augmented data in terms of the extended period from 1994:Q4 to 2019:Q4. Additionally, an extended wide sense replication is performed using an alternative variable measure of union power ($UnionR$), representing the actual union membership rate.

The results of the wide replication can reject the null hypothesis using both the F- and t- bounds test in favor of the existence of a cointegrating relationship (at a 5% level of significance) as in PSS. In the extended wide replication, the results suggested an even stronger favoritism of the existence of a level relationship, as every test rejected the null hypothesis.

A series of robustness analyses for the wide replication revealed that the long-run coefficients were stable over time and were not sensitive to the model choice (number of lags included). Further, by using long lags of unemployment, the speed of adjustment increases, probably due to the slow response of unemployment to changes in wages, as mentioned by Tyrvainen (1995).

The extended wide replication model (using $UnionR$) shows an even higher speed of adjustment ($\pi_y = -0.175$) without including unemployment lags.

The elasticities of productivity and unemployment are statistically significant, having the expected coefficients, with $\theta_{Prod} = 0.952$, thereby supporting the usual a priori hypothesis of the productivity coefficient being equal to unity and with $\theta_{UR} = -0.092$ being very close to what most other studies have found (i.e., a value around -0.1). The long-run coefficients of wedge and union membership are both statistically insignificant, meaning that no long-run effects exist. The estimate $\theta_{UnionR} = 0.042$ shows the lack of a long-run effect of union power on wages, and $\theta_{Wedge} = 0.107$ supports the hypothesis of a low real wage resistance.

Acknowledgements

We are grateful for the helpful and constructive comments of Professor Heather Anderson (Replication Section Editor) and three anonymous referees, which have significantly improved our paper. The usual disclaimer applies.

References

Arestis P, biefang-frisancho Mariscal I (1994). "Wage determiniation in the UK: further empirical results using conintegraion." *Applied Economics*, **26**(4), 417–424. doi:10.1080/00036849400000088.

Chan A, Savage D, Whittaker R (1995). "The new treasury model." *Government Economic Series Working Paper No. 128.* (Treasury Working Paper No. 70).

Cho JS, Greenwood-Nimmo M, Shin Y (2021). "Recent developments of the autoregressive distributed lag modelling framework." *Journal of Economic Surveys*, pp. 1–26. doi:https://doi.org/10.1111/joes.12450.

Dahl DB, Scott D, Roosen C, Magnusson A, Swinton J (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4, URL https://CRAN.R-project.org/package=xtable.

Darby J, Wren-Lewis S (1993). "Is there a cointegrating vector for UK wages?" *Journal of Economic Studies*, **20**, 87–115.

Hall SG (1986). "AN APPLICATION OF THE GRANGER & ENGLE TWO-STEP ESTIMATION PROCEDURE TO UNITED KINGDOM AGGREGATE WAGE DATA*." *Oxford Bulletin of Economics and Statistics*, **48**(3), 229–239. doi:https://doi.org/10.1111/j.1468-0084.1986.mp48003003.x.

Hebbali A (2020). *olsrr: Tools for Building OLS Regression Models*. R package version 0.5.3, URL https://CRAN.R-project.org/package=olsrr.

Johansen S (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, **59**(6), 1551–1580. ISSN 00129682, 14680262.

Kassambara A (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.4.0, URL https://CRAN.R-project.org/package=ggpubr.

Lupi C (2009). "Unit Root CADF Testing with R." *Journal of Statistical Software*, **32**(2), 1–19.

Natsiopoulos K, Tzeremes N (2021). *ARDL: ARDL, ECM and Bounds-Test for Cointegration*. R package version 0.1.1, URL https://CRAN.R-project.org/package=ARDL.

Nickell S, Andrews M (1983). "Real wages and employment in Britain." *Oxford Economic Papers*, **35**, 183–206.

Office for Budget Responsibility (OBR) (2013). "The macroeconomic model." *OBR Briefing paper No. 5.* URL https://obr.uk/docs/dlm_uploads/Final_Model_Documentation.pdf.

Peng RD (2011). "Reproducible research in computational science." *Science*, **334**, 1226–1227.

Pesaran MH, Shin Y (1999). *An Autoregressive Distributed-Lag Modelling Approach to Cointegration Analysis*, chapter 11. Econometric Society Monographs. Cambridge University Press.

Pesaran MH, Shin Y, Smith RJ (2001). "Bounds testing approaches to the analysis of level relationships." *Journal of Applied Econometrics*, **16**(3), 289–326. ISSN 08837252. doi:10.1002/jae.616. CAM.5093.

Qiu D (2015). *aTSA: Alternative Time Series Analysis*. R package version 3.1.2, URL https://CRAN.R-project.org/package=aTSA.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sax C, Eddelbuettel D (2018). "Seasonal Adjustment by X-13ARIMA-SEATS in R." *Journal of Statistical Software*, **87**(11), 1–17. doi:10.18637/jss.v087.i11.

Tang Y, Horikoshi M, Li W (2016). "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." *The R Journal*, **8**. URL https://journal.r-project.org/.

Trapletti A, Hornik K (2019). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47., URL https://CRAN.R-project.org/package=tseries.

Tyrvainen T (1995). "Real Wage Resistance and Unemployment: Multivariate Analysis of Cointegrating Relations in 10 OECD Countries." *OECD Jobs Study Working Papers*, (10). doi:https://doi.org/https://doi.org/10.1787/136050768537.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. URL https://ggplot2.tidyverse.org.

Zeileis A (2019). *dynlm: Dynamic Linear Regression*. R package version 0.3-6, URL https://CRAN.R-project.org/package=dynlm.

Zeileis A, Hothorn T (2002). "Diagnostic Checking in Regression Relationships." *R News*, **2**(3), 7–10. URL https://CRAN.R-project.org/doc/Rnews/.

Zeileis A, Leisch F, Hornik K, Kleiber C (2002). "strucchange: An R Package for Testing for Structural Change in Linear Regression Models." *Journal of Statistical Software*, **7**(2), 1–38. URL http://www.jstatsoft.org/v07/i02/.

Affiliation:

Kleanthis Natsiopoulos
Department of Economics, University of Thessaly
28th October street, 78, 38333, Volos, Greece
E-mail: klnatsio@econ.uth.gr
URL: https://natsiopoulos.netlify.app/

Nickolaos G. Tzeremes
Department of Economics, University of Thessaly
28th October street, 78, 38333, Volos, Greece
E-mail: bus9nt@econ.uth.gr
URL: https://sites.google.com/site/tzeremesn/