THREE ESSAYS ON POST–KEYNESIAN MODELS OF EFFECTIVE DEMAND AND INCOME DISTRIBUTION

by

Jose A. Barrales Ruiz

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Economics

The University of Utah

August 2017

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Jose A. Barrales Ruiz**
has been approved by the following supervisory committee members:

Rudiger von Arnim ,	Chair(s)	**14 Apr 2017** Date Approved
Korkut Erturk ,	Member	**14 Apr 2017** Date Approved
Richard Fowles ,	Member	**14 Apr 2017** Date Approved
Codrina von Arnim ,	Member	Date Approved
Christian Proaño ,	Member	Date Approved

by **Thomas Maloney** , Chair/Dean of
the Department/College/School of **Economics**
and by **David B. Kieda** , Dean of The Graduate School.

ABSTRACT

This dissertation consists of three independent essays on cyclical fluctuations of functional income distribution and effective demand in the Post–Keynesian tradition.

The first essay investigates the longer run relationship between wage share and measurements of economic activity. Our main tools are wavelet covariance and multiresolution analysis. Results indicate that (1) Goodwin type cycles are observed even at longer run and (2) when considering smooth trends for periodicities longer than 32 years, a *long* Goodwin cycle seems to appear from the 1940s to mid–1990s that collapses afterwards.

The second and third essays are related in the sense that they empirically investigate the possibility of strong internal dynamics in the business cycle model of effective demand and income distribution.

Specifically, in the second essay, we study wage share and output gap in an univariate setting. Each time series is examined through chaos theory. The main tools are the nonlinear autoregressive neural network model, the dominant Lyapunov exponent, coefficient of determination, and local Lyapunov exponent. Results indicate that output gap might behave quasi–chaotically and wage share noisy–stable.

Finally, the third essay inquires into the possibility of limit cycle in the two–dimensional model on wage share and output gap. For that, we use the multivariate nonlinear autoregressive neural network model. Our results indicate that limit cycle behavior describes well their dynamics and, furthermore, the instability is located on the wage share isocline.

Chapters 1 through 3 open several questions that we hope further research will address.

In Chapter 1, we conjecture that globalization plays a crucial role in the stagnation of the trends in the late 1990s. However, further research is required.

Chapter 2 concludes that instability is rooted in the goods market dynamics rather than the distributive dynamics.

Results on Chapter 3 indicate that the demand regime is stable and wage share locally unstable. This possibility remains largely unexplored both in the theoretical and empirical

literature, and it creates a contradiction with the results found in Chapter 2. Further research is necessary on the robustness of the result, and possible mechanisms.

To my wife, Pamela.

CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my wife, Pamela, who has been unconditional and followed me in this dream. I have always admired your heart and strength. Thank God for crossing our ways and give the opportunity to form this beautiful family. Also thanks to my family (in-laws too!) for all the support, love, and understanding.

I'd like also to thank my advisor, Rudiger (Rudi) von Arnim, who helped me to choose a topic that fitted best my skills. His constant support and guidance during seemly dead ends made me succeed in this enterprise.

Thanks to my family in Utah: Nathe, Nico, Jenny, and Pame. Thanks for your friendship. It is impossible to even summarize all the great moments we shared!

To my fellow comrades, Soumava and Miki, it was always great to share ideas and frustrations with you. Certainly, these years have been a lot more interesting and funnier with you. I really wish you the best in the future.

Last but not least, I want to thank God, overall, for putting all these great people in my way. All of you made this experience something I'll never forget.

CHAPTER 1

LONGER RUN DISTRIBUTIVE CYCLES: WAVELET DECOMPOSITIONS FOR THE US, 1948–2011

1.1 Abstract

This paper presents an analysis of the comovement of income–capital ratio, output gap and employment rate vis–á–vis the functional distribution of income. We decompose time series into wavelets of varying periodicity. Cycles at all periodicities in all three variables vis–á–vis wage share show a counter–clockwise ("Goodwin") pattern. The well–known regular cycles appear at business cycle frequency. Furthermore, one roughly 30–year cycle exists before 1980. Post–1980, no clear medium run cyclical picture emerges. This finding is complemented by wavelet covariance analysis, which suggests that covariance of longer period cycles is negative before 1980, but positive thereafter. Crucially, trajectories of trends across the entire postwar period raise the possibility of one "long" 60–year Goodwin cycle in all three variables vis–á–vis the wage share, which would suggest that sustained growth after 2000 required much broader real wage increases relative to labor productivity. We conduct non–parametric Granger tests, which indicate that systematic interaction at all periodicities exist. We discuss our findings in relation to the debate on wage–led and profit–led demand regimes.

1.2 Introduction

Ever since Rowthorn (1982), Dutt (1984), Taylor (1985, 1991), and Bhaduri and Marglin (1990), post–Keynesians of all stripes have debated interactions between demand and distribution under the labels of *wage–led* and *profit–led* demand. There are many different

entry points into this debate. Four issues in particular stand out.

The first is whether one approaches the interaction between demand and distribution as a (potentially locally unstable) cyclical process, or as a stable growth process (Barbosa-Filho & Taylor, 2006; Skott & Zipperer, 2012; von Arnim & Barrales, 2015). A related question is whether *both* variables are endogenous in either such specification, or not. Based on our analysis, we argue that treatment of the wage share as exogenous in empirical applications will lead to misspecification: results from non–parametric tests on Granger causality suggest that systematic "bi–directional" interaction exists at all periodicities between measures of economic activity and the functional distribution of income.

The second concerns the long and the short of it. Suppose for the moment that a consensus emerges in the literature that business cycles feature a profit–led and profit–squeeze "Goodwin" pattern (Barbosa-Filho & Taylor, 2006; Proaño, Flaschel, Ernst, & Semmler, 2009). Then the key research question concerns the interaction between growth and distribution in the longer run. Kiefer and Rada (2015) raise the possibility of long–run stagnation through labor suppression in a globalized economy, though all countries' business cycles are of "Goodwin type." Blecker (2016) discusses these and other findings from a theoretical perspective, suggesting that indeed the consumption or wage–related constraints to growth might be most relevant at longer time horizons. Our analysis here complements existing research on the possible existence of a *long* Goodwin cycle (Flaschel, 2010; Tavani, Flaschel, & Taylor, 2011), which raises questions about the mechanisms underlying stagnationary tendencies.

The third issue pertains to country coverage: should one analyze interaction of demand and distribution in a "closed system" or an open economy? The former would aim at the trade–off between consumption and investment, whereas the latter must take the trade channel and the possibility of a fallacy of composition at the global level into account (Blecker, 1989; Rezai, 2015; Sasaki & Fujita, 2015). While we focus here on the US, we see this issue as related to the previous point: any country might boost demand and employment through competitive devaluations. Analogously, a country might boost (investment) demand and employment in the short run through labor suppression. However, ultimately, capacity expansion must be justified, and effective demand must be supported by wages *somewhere*, and *at some point in time*.

The fourth deals with the endogeneity (or lack thereof) of the rate of utilization (Nikiforos, 2013, 2016; Skott, 2012). We will sidestep this question to some extent. It certainly matters, but our primary interest here is to document the existence of longer–run distributive cycles. We do not provide discussion of or empirical evidence in favor of a particular mechanism that attracts or repels firms to a steady state. It is however absolutely paramount to recognize that all relevant time series have significant medium or even long–term trends. Theory around Kaldor's stylized facts is one thing, and taking account of these longer trends another: business cycles are of Goodwin type (Mohun & Veneziani, 2008; Zipperer & Skott, 2011), but there is more to be said about the lower frequencies.

Thus, our intent is to document these lower frequencies. To do so, we present wavelet decompositions of income–capital ratio, output gap, and employment rate as well as the functional distribution of income for the US economy after World War II. The wavelet methodology decomposes a time series into cyclical movements with differing periodicities. It is not very widespread in economics. Gallegati and Semmler (2014) present various applications, and Gallegati, Gallegati, Ramsey, and Semmler (2011) provide a relevant discussion of Phillips curve dynamics. Bridji and Charpe (2015) employ similar methods to empirically analyze long–run interactions of distribution and growth in US, UK, and France. Their findings complement ours. The wavelet decompositions might thus shed new light on a subset of the issues outlined above. Indeed, we find quite regular cyclical comovement at all periodicities. Crucially, the vast majority of these cycles of the wavelet decompositions conform to the pattern first suggested in Goodwin (1967).[1]

Our contribution is to provide new evidence on distributive cycles over longer time horizons in the US economy, and relate this evidence to the debate on the nature of demand regimes as well as the recent crisis and its aftermath. Key findings are: cycles at all periodicities in all three variables vis–á–vis the wage share show a counter–clockwise Goodwin pattern. The well–known regular cycles appear at business cycle frequency. Furthermore, one roughly 30–year cycle exists before 1980. Post–1980, no clear medium run cyclical picture emerges. This finding is complemented by wavelet covariance anal-

[1] Goodwin's original and well–known contribution provided a parsimonious description of *classical* growth and social conflict in employment–wage share space, with a constant income–capital ratio. The investigation in this paper builds on more recent literature that concerns empirical work on such cycles.

ysis, which suggests that covariance of longer period cycles is negative before 1980, but positive thereafter.

The positive covariance post–1980 must be interpreted in light of the tremendous increase in income (and wage) inequality over the last 35 years. Trajectories of trends across the entire postwar period raise the possibility of one "long" 60–year Goodwin cycle in all three variables vis–á–vis the wage share. That long cycle, however, appears to have been interrupted in the late 1990s, as the institutions that could have supported a phase of prosperity had been too thoroughly dismantled. If this interpretation is correct—and it is obviously hard to make inferences based on *one* long cycle—the popular interpretation of labor suppression and inequality as a cause of crisis and stagnation would hold water, but would as well need to be seen within the context of a long run profit–led growth cycle.

Additionally, demand and distribution "Granger–cause" each other at all periodicities. This result implies that the distribution of income cannot be considered an exogenous variable that explains demand, as has been done in some empirical studies (Hein & Vogel, 2008; Onaran & Galanis, 2012).

Next, we discuss the raw data. In subsequent sections, we briefly describe the wavelet methodology, provide wavelet decompositions and present the resulting cycles, and discuss.

1.3 Demand, employment, and distribution

Post–Keynesian research on growth and distribution usually considers the rate of utilization to be the relevant measure of demand. The rate of utilization is the ratio of observed output to "full capacity output," that is, $U \equiv Y/Y^*$. The problem is, of course, that the latter is unobservable. We do not know what "full capacity output" is, and even if we knew what it just had been, it might by now have changed. Further, it is not clear whether the ratio of actual to full capacity output is constant in the long run, as is whether that ratio converges over any time horizon to a "desired" rate of utilization (Nikiforos, 2013).

One proxy for the rate of utilization—for example, in Barbosa-Filho and Taylor (2006)—is the ratio of actual to Hodrick–Prescott (HP) trend output. The HP filtering method has a couple of crucial and well–known disadvantages. First, near the sample end, as the

filter runs out of observations, the trend tends to bend too much towards the latest cycle. Second, the resulting series of the rate of utilization has no more medium–term trends. This does not need to be problem, but it of course precludes analysis of the longer–term interactions between demand and distribution: the focus must be on the short run. Thus, here we employ two different measures. First, we use the *income–capital ratio*, $u \equiv Y/K$. It is an obvious candidate, since this ratio and the rate of utilization are proportional, as long as the full capacity output to capital ratio—$\sigma \equiv Y^*/K$—is constant. Then, $u \equiv U\bar{\sigma}$. In a messy real world, that does of course need not be the case. Further, in the long run, the ratio Y/K might be interpreted as "capital productivity," and would then not be a measure of demand. Coincidentally, of course, a constant long run u would imply steady growth. We will sidestep such theoretical complications for the moment, as our primary interest is to document that relevant longer–run movements in Y/K exist.

We construct the series as the ratio of corporate business (nominal) net value added to corporate business (nominal) net fixed assets at replacement costs. Net value added of corporate business is obtained from U.S. Bureau of Economic Analysis (BEA) National Income and Product Accounts Tables (NIPA) table 1.14, and is available as a quarterly series between 1948:I and 2014:IV.[2] Using the 1947 year–end estimate as a starting point, these observations are interpolated to generate a quarterly series. As a result, quarter–to–quarter cyclicality is driven by the numerator. However, one might assume that the relative smoothness of the denominator corresponds to the more gradual installment and implementation of, say, a purchased machine. The top left panel of Figure 1.1 shows the income–capital ratio and its trend against The National Bureau of Economic Research (NBER) recession dates. (The dashed "trend" line represents cyclical movements with period length greater than $2^7 = 128$ quarters per cycle. See section 1.4 for further detail.) As is already clear from this picture, the series exhibits cycles at different periodicities. The "Great Society" project and escalation of the Vietnam War in the 1960s correspond with a postwar high of the series, followed a distant second by the peak of the new economy bubble. These peaks can be contrasted with the longer–run cyclical troughs of

[2]Net fixed assets of corporate business are available as an annual series of year–end estimates in BEA NIPA table 6.1. Basu and Vasudevan (2013) use the ratio of the annual series as a measure of "capital productivity." See, for example, their Figure 17, p. 78 and related discussion.

the Volcker–Reagan recession and the global crisis of 2008. The latter marks the postwar low.

Second, we construct a proxy for demand as the ratio of Gross Domestic Product (GDP) to the Congressional Budget Office's (CBO) estimate of "potential output." The CBO's series of potential output is estimated based on a neoclassical production function framework. In so many words, the key assumption is to fit a measure of the total labor supply, rather than those actually in employment, in a growth accounting framework. The method generates a markedly different estimate of "full employment output" than, say, the HP filter in the aftermath of the crisis of 2008—exactly because it does not put undue weight onto recent observations of output.

We label this the *output gap* series. The CBO provides real potential GDP in chained 2009 dollars, so that we use the corresponding measure for real GDP in chained 2009 dollars (BEA NIPA table 1.1.6). Quarterly series are available between 1949:I and 2014:IV. The top right panel of Figure 1.1 presents the ratio and its trend against shaded NBER recessions. The evolution of the output gap is roughly similar to the income–capital ratio, but its peaks and troughs are more evenly distributed: the late 1960's stand less out, and the troughs of (roughly) 1982 and 2008 are of comparable depths. Still, the longer–term movement of these first two proxies for the rate of utilization is similar. The bottom right panel shows, for the sake of comparison, the ratio of GDP to its HP trend, with a standard smoothing parameter of 1600 for quarterly data. As can be seen, this proxy for the rate of utilization vastly underestimates the most recent contraction, and suggests that the US economy is well on its way of recovery. Especially the capital–income ratio and the output gap series show quite different pictures.

Now let us move on to the employment rate, shown in the bottom left panel of Figure 1.1. Following Tavani et al. (2011) and Zipperer and Skott (2011), we measure the employment rate as the remainder to unemployment: $e \equiv 1 - ur$, where the latter is the standard BLS measure of the civilian unemployment rate. The unemployment rate abstracts from the strong trends in the employment–population ratio of prime–age working adults, which reflects the strong increase in the female participation rate. Nevertheless, this measure exhibits the well–known medium term trends, and the cyclicality of shorter periods is quite the same.

Last but not least, let us consider the functional distribution of income. We focus here on *corporate* labor costs as a share of corporate net value added. To begin, look at the top–left panel of Figure 1.2. (The data underlying the figure stems from BEA NIPA table 1.14, and are available at quarterly frequency between 1948:I and 2014:IV.) The picture shows the ratio of compensation of employees to net value added, which is analogous to the calculation of the profit share in Zipperer and Skott (2011). As can be seen in the top–left panel, the series peaks around 2000, at the end of the "new economy" boom. Clearly, the measure does not show the suppression of labor associated with the second era of globalization (since 1980). It should therefore be emphasized that the shift of the postwar peak of the wage share from roughly 1980 to 2000 is due to recent data revisions of the BEA: the dot–dashed line show the data until 2013:I, the last revision where the peak of 1980 is still larger than 2000.[3]

To illustrate, consider the components of net value added, in the right panel of the middle row of Figure 1.2. Compensation of employees contains wages and salaries and supplements. The latter include employer contributions to pension funds, health insurance, and social insurance. In 2003, supplements made up about 12 percent of NVA; of these, about one–third were paid into public programs, and two–thirds into private funds.

Should these supplements be accounted for as part of the wage share? Due to the relative constancy of the share of production taxes in NVA, the answer to this question matters for the evolution of the wage share. The top–right panel shows wages and salaries as a share of NVA. This measure has a strong downward trend through the entire postwar period. Standard narratives—"profit–squeeze" during the *Golden Age* leading to social conflict and inflation in the 1970s, and restoration of profitability through conservative policy since roughly 1980—appear to depend to a very large degree on these supplements.[4]

One can conceive of arguments for either position. These flows are not part of take–home pay, are not wages, and do not directly enter the income stream. They act, rather,

---

[3]Details on these revisions are available in the BEA's NIPA tables, and from the authors upon request.

[4]Further, the dash–dotted line in the top right panel shows the BLS' "headline" non–farm business labor share, which adds wages and salaries to the portion of proprietor's income that the statisticians believe to be wage payments for self–employment. Crucially, for our purposes, only the second half of the Aughts appears to differ. See Bridji and Charpe (2015) for further discussion of measures of the wage share that include adjustments for self–employment.

as redistributive stabilizers, either across the labor force or across generations. Moreover, the distribution of benefits across the wage and salary structure is possibly skewed to the top. On the other hand, it might be argued that supplements should be included in the wage share, since these flows accrue to people who obtain an *overwhelming* majority of their lifetime income in wages. Feldstein (2008) makes this case, arguing that "[b]ecause of the rise in fringe benefits and other noncash payments, wages have not risen as rapidly as total compensation. It is important therefore to [consider] total compensation rather than [...] the narrower measure of just wages and salaries." As we are considering the wage share as an indication of real *labor cost*, this would seem reasonable. Moreover, the wage share provides the funds to support consumption, and a high share of these benefits does represent consumption expenditures. Indeed, one—quite classical—way to think of the wage share is to consider it the portion of net product that sustains the labor force. In other words, it is the portion of net product necessary to reproduce labor power. Health and retirement benefits in a modern capitalist economy should be considered part of that.

Still, these measures of the wage share do not show the secular rise and fall of real wages relative to labor productivity throughout postwar history. Rapidly rising wage inequality has masked these developments—and the very top of the wage income distribution should not be considered necessary consumption. We therefore adjust the *compensation* share of net value added to exclude the top 1% of the wage distribution (Piketty & Saez, 2003).[5] In the post–Keynesian literature, Tavani and Vasudevan (2014), for example, motivate their theoretical exploration of worker–manager conflict with these data. Relatedly, Carvalho and Rezai (2016) show that the top quintile of the wage income distribution already has a high savings rate of 40 percent. The very top of the wage distribution thus is a lot more "like" capital income, and in what follows we focus on this adjusted wage share: as can be seen, the peak of this series (middle left panel) is around 1980, and exhibits a sustained rise and fall before and after that.

The selection of the appropriate wage share measure is most important for the trend

[5]To be perfectly precise, this adjusted wage share is $(1-p)*(C/NVA)$, where p is the top 1 percentile of the wage income distribution, and C/NVA is the share of compensation, which includes supplements, relative to corporate net value added (top–left of Figure 1.2). We obtain a quarterly value for p by interpolating the annual series from Piketty and Saez (2003), Table B.2. Combining data from different sources in this fashion is not unproblematic, but for the lack of alternatives, we use this shortcut.

over the entire postwar period. However, the bottom row of Figure 1.2 illustrates that it does as well matter for longer–run cycles. Of the three different 16–32 year cycles shown in the bottom–right panel, only the (gray) adjusted wage share has a lower peak in 2000 than in 1980. For cycles with shorter periodicities, shown in the left panel, the question is secondary.

In the following section we discuss the wavelet methodology in more detail.

1.4 Empirical methods

Our empirical methods are wavelet transform and multiresolution analysis (MRA), as well as a non–parametric Granger causality test. In this section we describe them in turn.

1.4.1 Wavelet transform and multiresolution analysis

A complete discussion of wavelet methods for time series can be found, for instance, in Percival and Walden (2000). First, it should be noted that economic processes unfold at different time scales. The discrete wavelet transform (DWT) provides a natural way to decompose a time series in a multiresolution fashion that allows analysis of the time series' variance at these different time scales.

We employ the maximal–overlap DWT (MODWT). The MODWT for level J for a time series $\boldsymbol{X}$ yields highly redundant and non–orthogonal column vectors $\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \ldots, \tilde{\boldsymbol{W}}_J$ and $\tilde{\boldsymbol{V}}_J$ each of dimension N. The vector $\tilde{\boldsymbol{W}}_j$ contains the so–called wavelet coefficients and is associated with changes in $\boldsymbol{X}$ on a scale of $\lambda_j \Delta t$, with $\lambda_j = 2^{j-1}$, while $\tilde{\boldsymbol{V}}_j$ are called scaling coefficients and are associated with variations at scales $\lambda_{J+1}\Delta t$. The wavelet and scaling coefficients are associated with zero phase filters, which improves alignment with the original series.

MODWT yields an energy decomposition

$$||\boldsymbol{X}||^2 = \sum_{j=1}^{J} ||\tilde{\boldsymbol{W}}_j||^2 + ||\tilde{\boldsymbol{V}}_J||^2, \tag{1.1}$$

where $||.||$ is the l^2–norm, and an additive decomposition called multiresolution analysis (MRA)

$$\boldsymbol{X} = \sum_{j=1}^{J} \tilde{\boldsymbol{\mathcal{D}}}_j + \tilde{\boldsymbol{\mathcal{S}}}_J, \tag{1.2}$$

where $\tilde{\mathcal{D}}_j$ and $\tilde{\mathcal{S}}_J$ are the j–th order *detail* and the J–th order *smooth* for $\boldsymbol{X}$ (see Percival & Mofjeld, 1997, p. 3). Each wavelet and scaling coefficient are obtained by

$$\tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N} \quad \text{and} \quad \tilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{t-l \bmod N} \tag{1.3}$$

with $t = 0, \ldots, N-1$. The coefficients $\{\tilde{h}_{j,t}\}$ and $\{\tilde{g}_{j,t}\}$ are called wavelet and scaling filters with width $L_j = (2^j - 1)(L-1) + 1$. For instance, for the Haar wavelet the wavelet coefficients are $\tilde{h}_{1,0} = 0.5$ and $\tilde{h}_{1,1} = -0.5$ and scaling coefficients $\tilde{g}_{1,0} = 0.5$ and $\tilde{g}_{1,1} = 0.5$.

In matrix notation, the transform from $\boldsymbol{X}$ to $\tilde{\boldsymbol{W}}_j$ and from $\boldsymbol{X}$ to $\tilde{\boldsymbol{V}}_j$ can be expressed as

$$\tilde{\boldsymbol{W}}_j = \tilde{\mathcal{W}}_j \boldsymbol{X} \quad \text{and} \quad \tilde{\boldsymbol{V}}_j = \tilde{\mathcal{V}}_j \boldsymbol{X}, \tag{1.4}$$

where each row of the $N \times N$ matrix $\tilde{\mathcal{W}}_j$ has a value of $\{\tilde{h}^\circ_{j,t}\}$ and $\tilde{\mathcal{V}}_j$ of $\{\tilde{g}^\circ_{j,t}\}$, where filters $\{\tilde{h}^\circ_{j,t}\}$ and $\{\tilde{g}^\circ_{j,t}\}$ are $\{\tilde{h}_{j,t}\}$ and $\{\tilde{g}_{j,t}\}$ periodized to length N, respectively. The MRA is obtained, therefore, as

$$\tilde{\boldsymbol{\mathcal{D}}}_j \equiv \tilde{\mathcal{W}}_j^T \tilde{\boldsymbol{W}}_j \quad \text{and} \quad \tilde{\boldsymbol{\mathcal{S}}}_j \equiv \tilde{\mathcal{V}}_j^T \tilde{\boldsymbol{V}}_j, \tag{1.5}$$

so that we can write in analogy to equations (1.3)

$$\tilde{\mathcal{D}}_{j,t} = \sum_{l=0}^{N-1} \tilde{h}^\circ_{j,l} \tilde{\boldsymbol{W}}_{j,t+l \bmod N} \quad \text{and} \quad \tilde{\mathcal{S}}_{j,t} = \sum_{l=0}^{N-1} \tilde{g}^\circ_{j,l} \tilde{\boldsymbol{V}}_{j,t+l \bmod N}. \tag{1.6}$$

In our application, the main objective is to analyze variance and covariance at time scales that describe the longer run. The Haar wavelet filter is appropriate in these circumstances. It does not provide the best approximation to the band–pass filter, since the number of filter coefficients is much smaller. However for the same reason the number of boundary coefficients is the smallest.

This small number of boundary coefficients, that is, those wavelet coefficients calculated using the boundary assumption, permits calculation of an unbiased estimator for variance and covariance at longer scales, which is not possible with filters that require more coefficients. The computation of the wavelet coefficients uses the boundary assumption of circularity, which is well situated when the starting point is close to the last point in the series. Since this is usually not the case, we made use of the reflexive boundary condition, which attaches the reverse series at the end of the original one.[6]

[6] A detailed discussion of wavelet variance and covariance analysis and their confidence intervals can be found in Percival (1995), Percival and Walden (2000), and Whitcher, Guttorp, and Percival (1999, 2000).

Specifically, the unbiased estimator for wavelet variance with scale coefficient $\lambda_j = 2^{j-1}$ is calculated as

$$\hat{\nu}^2(\lambda_j) = \frac{1}{N_j} \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,t}^2 \tag{1.7}$$

where $N_j = N - L_j + 1$ is the number of non–boundary coefficients at the jth–level, with $L_j = (2^j - 1)(L-1) + 1$ and L is the number of filter coefficients. For the Haar wavelet filter, $L = 2$. Percival and Walden (2000, p. 312), show that under appropriate conditions an approximate $100(1-\alpha)\%$ confidence interval for $\hat{\nu}^2(\lambda_j)$ is given by

$$\left[\frac{\eta_j \hat{\nu}^2(\lambda_j)}{Q_{\eta_j}(1-\alpha/2)}, \frac{\eta_j \hat{\nu}^2(\lambda_j)}{Q_{\eta_j}(\alpha/2)} \right], \tag{1.8}$$

where $Q_{\eta_j}(\alpha)$ is the $\alpha \times 100\%$ percentage point for the χ^2 distribution with η_j equivalent degrees of freedom. Further,

$$\begin{aligned} \hat{\eta}_j &= \frac{N_j \hat{\nu}^4(\lambda_j)}{\hat{A}_j} \text{ and} \\ \hat{A}_j &= \frac{\hat{\nu}^4(\lambda_j)}{2} + \sum_{\tau=1}^{N_j-1} \left(s_{j,\tau}\right)^2 \end{aligned}$$

where $s_{j,\tau}$ is the biased estimator of the autocovariance sequence (ACVS) and A_j the integral of the spectral density function for $\tilde{\boldsymbol{W}}_j$. The unbiased estimator of the wavelet covariance for series X and Y is given by

$$\hat{\gamma}_{XY}(\lambda_j) = \frac{1}{N_j} \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,t,X} \tilde{W}_{j,t,Y}. \tag{1.9}$$

The method for constructing its confidence interval is described in detail in Whitcher et al. (1999, 2000). The approximate $100(1-\alpha)\%$ confidence interval for $\gamma_{XY}(\lambda_j)$ is

$$\left[\hat{\gamma}_{XY}(\lambda_j) - \Phi^{-1}(1-\alpha/2) \sqrt{\frac{\tilde{\sigma}_{j,XY}^2}{N_j}}, \hat{\gamma}_{XY}(\lambda_j) + \Phi^{-1}(1-\alpha/2) \sqrt{\frac{\tilde{\sigma}_{j,XY}^2}{N_j}} \right] \tag{1.10}$$

where

$$\begin{aligned} \tilde{\sigma}_{j,XY}^2 &\equiv \frac{\hat{s}_{j,0,X}\hat{s}_{j,0,Y}}{2} + \sum_{\tau=1}^{N_j-1} \hat{s}_{j,\tau,X}\hat{s}_{j,0,Y} + \frac{1}{2} \sum_{\tau=-(N_j-1)}^{N_j-1} \hat{s}_{j,\tau,XY}^2 \text{ with} \\ \hat{s}_{j,\tau,X} &= \frac{1}{N_j} \sum_{t=L_j-1}^{N-|\tau|-1} \tilde{W}_{j,t,X} \tilde{W}_{j,t+|\tau|,X} \text{ and} \\ \hat{s}_{j,\tau,XY}^2 &= \frac{1}{N_j} \sum_t \tilde{W}_{j,t,X} \tilde{W}_{j,t+\tau,Y} \end{aligned}$$

and the summation is over $t = L_j - 1, \ldots, N - \tau - 1$ for $\tau \geq 0$ and over $t = L_j - \tau - 1, \ldots, N - 1$ for $\tau < 0$.

1.4.2 Non–parametric Granger causality

In order to test for Granger causality, we use the non–parametric test proposed by Diks and Panchenko (2006) (DP test) instead of Hiemstra and Jones (1994)'s modified Baeck and Brock (1992) non–parametric test, which has been reported to over reject the null hypothesis.

Granger causality test seeks to detect evidence against the null hypothesis that $\{X_t\}$ does not "Granger–cause" $\{Y_t\}$. In terms of their probability density, the null hypothesis can be written as follow:

$$Y_{t+1} \left| \left(X_t^{l_x}, Y_t^{l_y} \right) \right. \sim Y_{t+1} \left| Y_t^{l_y} \right. \tag{1.11}$$

where $X_t^{l_x} = (X_{t-l_x+1} + \cdots + X_t)$ and $Y_t^{l_y} = (Y_{t-l_y+1} + \cdots + Y_t)$, and "$\sim$" denote equivalence in distribution. Consider the $(l_x + l_y + 1)$–dimensional vector $W_t = (X_t^{l_x}, Y_t^{l_y}, Z_t)$ with $Z_t = Y_{t+1}$. Dropping the time index for simplicity, the null hypothesis is a statement about the equality of

$$\frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)} \frac{f_{Y,Z}(y,z)}{f_Y(y)} \tag{1.12}$$

The DP test statistics then can be written as

$$T_n(\varepsilon) = \frac{(n-1)}{n(n-2)} \sum_i \left(\hat{f}_{X,Y,Z}\left(\boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \hat{f}_Y\left(\boldsymbol{Y}_i\right) - \hat{f}_{X,Y}\left(\boldsymbol{X}_i, \boldsymbol{Y}_i\right) \hat{f}_{Y,Z}\left(\boldsymbol{Y}_i, \boldsymbol{Z}_i\right) \right) \tag{1.13}$$

with

$$\hat{f}_W\left(\boldsymbol{W}_i\right) = \frac{(2\varepsilon)^{-d_W}}{n-1} \sum_{j, j \neq i} \boldsymbol{I}_{ij}^W \quad \text{and} \quad \boldsymbol{I}_{ij}^W = \boldsymbol{I}(||\boldsymbol{W}_i - \boldsymbol{W}_j|| < \varepsilon)$$

with $||.||$ the maximum norm. Diks and Panchenko (2006, Theorem 1 on p. 1656) provides a proof of the asymptotic distribution of the estimator.

Typically, the test is executed as a two–stage procedure: first, one estimates a VAR model which filters out the linear relationship. Second, one conducts the DP test on the standardized residual error terms, where only the nonlinear behavior remains. The main problem here is the dependence on this previous step which can either withdraw too much nonlinear information or too little linear information—leaving the residuals not informative enough. Moreover, Pitarakis (2006) shows, in the context of threshold

autoregressive models, that usual lag length selection methods (BIC, AIC, etc) fail if the true model is highly nonlinear. This would potentially distort the testing.

For that reason we build on the proposal of Li and Shukur (2011). They test for non-linear Granger causality under a logistic smooth transition autoregressive (LSTAR) model, using the test in Hiemstra and Jones (1994) over the J–order smooth, that is, $\tilde{\mathbf{S}}_J$. We use the DP test over the *standardized* j-order details, that is, $\tilde{\mathbf{D}}_j$. Lastly, we set $l_x = l_y = L_j$, as well as $\varepsilon = 1.5$, which is the recommended level given the number of observations.

1.5 Results and discussion

In this section, we present and discuss empirical results. Our analysis suggests that the well–known profit–led/profit–squeeze business cycles are embedded in a long—60–year—Goodwin pattern of all three activity measures against the functional distribution of income. The long cycle, however, appears to have been disrupted in the late 1990s. A sustained phase of prosperity with rising demand, employment, and real wages relative to labor productivity should have occurred.

1.5.1 From short run to 30–year cycles

To get started, we provide an example of the multi–resolution analysis, namely the reconstructed time series at four different time scales for output gap and wage share. Figure 1.3 shows these four panels. From top left to bottom right, the panels present business cycles with period length between 2–4 and 4–8 years as well as longer cycles with period length from 8–16 and 16–32 years. For an example, consider the 8–16 year cycles in the bottom–left panel. These cycles pick up the immediate postwar expansion, the upswing in the 1960s, Reagan's public deficit, the new economy boom, and the 1970s and 2000s in the doldrums. What is already apparent in this picture is the *sequence* of movements: demand increases from its trough while the wage share is decreasing. Only once the demand expansion is well on its way does the wage share turn upward. This is, of course, the "Goodwin" cycle, and we will return to it in just a moment.

The bottom right–panel of the figure shows the longest period length, lasting from 16 to 32 years. Here the wavelet decomposition shows cyclical swings that clearly extend across several business cycles. The output gap series shows three long cycles: the postwar

expansion, the *Great Society* expansion, and the Reagan–Clinton era of hyperglobalization. The two sustained downturns cover the 1970s and the 2000s. The recovery since the crisis of 2008 is substantially incomplete. The corresponding movements of the wage share are less pronounced. At all period lengths, the wage share shows a cyclical trough—as in the original series—in the mid–1960s, and cyclical highs near that. The 16–32 year series collapses after 2000, and has not recovered.

Similar charts could be put together for output gap and employment rate vis–à–vis the wage share. The focus here, however, is on the cyclical patterns, so that we present the data as such. To begin, we illustrate the Goodwin pattern at business cycle frequency. Figure 1.4 shows wavelet decompositions of output gap and wage share against each other in an empirical phase diagram. (The data are the same as at the top–right of Figure 1.3.) These cycles correspond to 4–8 year cycles. The top–left panel begins in 1949, and the bottom–right ends in 2011. The direction of the cycles can be traced easily: the dot in each panel marks the beginning. Two main observations stand out. First, these nine panels correspond quite remarkably well with the NBER recession dates—which would suggest ten recessions in the postwar era. Second, every single panel portrays a Goodwin pattern, with the seemingly clockwise swerve in the early 1990s the exception. Careful econometrics on this or similar data describing the business cycle fluctuations around a trend will tend to pick up a profit–led and profit–squeeze mechanism. At this frequency, the Goodwin pattern is fully apparent even in the latest cycle of the Aughts.

However, in what follows, the focus will be on cycles with longer frequency. Figure 1.5 shows phase trajectories of income–capital ratio (on the horizontal axis) and the wage share, Figure 1.6 shows phase trajectories of output gap and wage share, and Figure 1.7 shows employment rate and wage share. Each of these show 8–16 year cycles in the top row, and 16–32 year cycles in the bottom. The left– and right–hand panels show the sub–samples pre and post 1980, as that roughly splits the sample in half—but additionally marks the turning point from Golden Age and social conflict to hyperglobalization. The 8–16 year series show—roughly—five cycles for the US postwar period. That is half the number of NBER cycles. However, the average period length is about 12 years—rendering these fluctuations still broadly in sync with "longer" business cycle expansions and contractions. Importantly, all cycles at this frequency show a Goodwin pattern, with the

exception of the clockwise nook of the income–capital ratio in the late 1950s (see top–left panel of Figure 1.5).

The two bottom panels in these three figures display 16–32 year cycles. As is apparent across all three, the period before 1980 and after 1980 differ drastically. Income–capital ratio, output gap, and employment rate show a complete counter–clockwise cycle from the mid–1950s to 1980. The few years before that do not strike one as contradictory to that pattern. Moreover, these pre–1980 30–year cycles are *tilted*, such that the economy is situated mostly in areas where one variable increases and the other decreases. (We return to that in a moment.) However, post–1980, the activity measures initially increase, though without a sustained rise in the wage share, and in the year 2000 (income–capital ratio) or 2001 (output gap and employment rate) collapse in lockstep with the distribution of income.

To further investigate this pattern, we present an analysis of the covariance of the three variables against the wage share. See Figure 1.8 for an overview, and section 1.4.1 for details on the calculations. In each panel, the horizontal axis shows the time scale of the wavelet decomposition, from shortest (0.5–1 year, time scale 1) to longest (16–32 years, time scale 6). The vertical axis represents the covariance; the dashed lines are confidence intervals. Let us dissect the top row for the income–capital ratio in detail. The leftmost panel shows the entire sample period. The covariance decreases from time scale 1 to 4, and then increases slightly. This, however, masks the sharp differences across periods. The middle panel illustrates that the covariance decreases across time scales 1–6 before 1980—but after 1980, it decreases to scale 4 and then becomes positive. This pattern can be observed in similar fashion for the output gap and employment rate. Thus, roughly speaking, before 1980, all time scales show negative covariance, which corresponds at scale 6 to the *tilted* cycles discussed above. But post–1980, the covariance at that scale turns positive, even if the business cycle covariance (recall Figure 1.4) remains negative.[7]

Such covariance analysis can not substitute for further empirical research on the mechanisms that drive the observed trajectories. However, the phase trajectories across business

[7]One possible interpretation of the negative covariance draws on the familiar phase diagram: the economy is situated on average *nearer* the goods market equilibrium (the profit–led demand isocline) than the labor market equilibrium (the profit–squeeze distributive isocline). While the goods market might be on average in disequilibrium, it nevertheless adjusts faster than the labor market.

cycles (Figure 1.4) and those across longer time horizons (Figures 1.5, 1.6, and 1.7) suggest that at all frequencies, activity measures, and the functional distribution of income *interact systematically*. To check whether these interactions are statistically significant, we conduct tests on Granger causality; see section 1.4.2 for details. Table 1.1 summarizes results. Let us focus here on time scales 3 through 6.

Statistically, the linkages between the pairs of variables are strongest—and always bi–directional—at scale 3. The 4–8 year interaction is bi–directional and strong for output gap and employment rate, but would suggest that the income–capital ratio does not Granger–cause the wage share. Similarly, at the lowest frequencies, the wage share Granger–causes income–capital ratio, output gap, and employment rate, but not always—and not always as strongly—vice versa. However, across all three pairs, the evidence indicates that statistically significant systematic interaction between the functional distribution of income and measures of economic activity exists. Such evidence does not support a specific mechanism, nor does it suggest relevant underlying ultimate causes that are potentially driving longer run cycles. Skott (2017) discusses these issues, specifically that research should focus on changes in exogenous factors, and try to delineate how shifts in these affect *both* growth and distribution. The Granger exercise here does not speak to these questions, but it highlights the possibility that regressions of the wage share on (components of) demand are likely misspecified.

Moreover, the covariance analysis emphasizes that something might be rotten with the demand regime: *if* the pre–1980 30–year cycle is representative of capitalist interaction between growth and distribution, then the subsequent 30–years would have to be seen as an aberration. In this sense, the positive covariance post–1980 is the manifestation of the collapse of activity measures and wage share in tandem. Of course, the well–known trends of financialization and globalization come to mind, and one might conjecture: demand and employment could indeed be sustained through reorganization of global production processes, and the related innovation and investment, *but not forever*.

1.5.2 A long wave: A 60–year Goodwin cycle?

To further illustrate, Figure 1.9 displays the *longest* possible cycle—the comovement of the trend of income–capital ratio, output gap, and employment rate vis–á–vis the trend

of wage share. The left panels display cycles with period length greater than $2^7 = 128$ quarters, or 32 years.[8] In the panels on the right, we define for the purpose of comparison the trend as the components of the series with periodicity greater than $2^6 = 64$ quarters, or 16 years. As a result, the right panels show slightly more cyclicality than the left panels.

The charts suggest that the postwar period might be marked by *one* long Goodwin cycle in all three activity measures and the wage share between, roughly, the early 1950s and the late 1990s. At the end of this period, economic activity and the wage share collapse together. The turning points of these trajectories roughly sketch a possible interpretation of postwar US developments: the Golden Age turns into a secular profit–squeeze in the mid–1960s; at about 1980 (or 1982), the conservative reaction breaks the upward trend in the wage share, and engineers capitalist revival, driven by increasing profitability. Now, what does the south–west–ward collapse of wage share and activity in the late 1990s represent?

At least one possibility arises. For the sake of argument, suppose we knew with certainty that a "long" Goodwin cycle exists. Then the wage share should have begun to rise in the late 1990s, triggering a *phase of prosperity*.[9] During this phase of prosperity, real wages relative to labor productivity should have been rising, demand relative to capacity should have been increasing, the output gap should have been shrinking, and the employment rate should have been rising. Crucially, such sustained increases in real incomes (and related consumption) might have justified innovation, capacity expansion, and global reorganization of production processes. Figure 1.10 provides a different look at the relevant period. We overlay the smooth trend with 4–8 year business cycles. The visualization raises the very possibility that the long cycle indeed could have been completed. Instead, what happened was Clinton, and Bush, and hyperglobalization, and ultimately Bear Stearns and Lehmann. In the frenzy over technology and finance, the real economy was forgotten, and it broke.

What are the implications for the debates on demand regimes? For the sake of argument, again, suppose that there is a long Goodwin cycle. Growth across this cycle

---

[8] For the top left panel, these data are the same as the dashed trend lines in the top left panel of Figure 1.1 and the middle left panel of Figure 1.2.

[9] We borrow the phrase from Charpe and Proaño (2016), who discuss the possibility that either wage–led or profit–led demand regimes could be embedded in a long Goodwin cycle; see their Figure 2.

would be *profit–led*, and the distributive regime would show a *profit–squeeze*. Nevertheless, growth and employment would at times—such as the late 1990s and 2000s—require higher real wages. The growth cycle will not keep turning, if real wages do not "overshoot." Erturk (2016) suggests a possible mechanism: globalization rendered labor everywhere weak, and capitalists unable to coordinate appropriate and broad real wage increases. A similar and well–known narrative applies to business cycles Barbosa-Filho and Taylor (2006). However, there is a crucial difference: over the length of the cycle, institutions can be taken as exogenous. In the long run, institutions evolve. Indeed, the trajectories in Figure 1.9 should probably be seen as driven by institutional changes, from the emergence and strengthening of "New Deal" institutions post World War II to its dismantling in the 15 years after the Reagan–Volcker shock.

Lastly, investigations on *steady* growth potentially miss the point, somewhat. Growth might be steady across several long Goodwin cycles—but across two or three business cycles, the process of growth and conflict does not evolve in a vacuum. Rather, it should possibly be evaluated in the context of a long cycle. It would follow that restoration of profitability 1980 became paramount; as it would follow that capitalist elites should have sought to engineer a long *wage supported* recovery, beginning 1995.

1.6 Conclusions

The literature has dissected demand regimes theoretically and empirically from a variety of viewpoints. All available evidence strongly suggests that at business cycle frequencies, measures of economic activity and functional distribution of income comove in a Goodwin pattern. Such a pattern features profit–led demand and profit–squeeze distribution. Most contradictory evidence relies on specifications that assume the distribution of income to be exogenous, which is hard to defend.

Additionally, longer–run distributive cycles exist not only in employment rate and functional distribution of income, but as well vis–á–vis income–capital ratio and output gap. The US trajectories suggest the possibility of an "aborted" long Goodwin cycle: instead of a phase of prosperity post–1995 with rising demand, employment, and real wages, all of these measures collapse together.

A possible culprit is globalization. How could any single country allow sustained real

wage increases? The pressures of international competition are immense. As such, the collapse of the long growth cycle represents a failure of elites to recognize the relevant linkages, and to maintain and support global institutions that can guide these processes in more sustainable ways.

1.7 References

Baeck, E. G., & Brock, W. A. (1992). *A general test for nonlinear granger causality: Bivariate model.* Working Paper, Iowa State University and University of Wisconsin, Madison, WI.

Barbosa-Filho, N., & Taylor, L. (2006). Distributive and demand cycles in the US economy – a structuralist Goodwin model. *Metroeconomica*, *57*(3), 389–411.

Basu, D., & Vasudevan, R. (2013). Technology, distribution and the rate of profit in the US economy: Understanding the current crisis. *Cambridge Journal of Economics*, *37*(1), 57-89.

Bhaduri, A., & Marglin, S. (1990). Unemployment and the real wage – the economic–basis for contesting political ideologies. *Cambridge Journal of Economics*, *14*(4), 375-393.

Blecker, R. (1989). International competition, income–distribution and economic–growth. *Cambridge Journal of Economics*, *13*(3), 395-412.

Blecker, R. (2016). Wage–led versus profit–led demand regimes: The long and the short of it. *Review of Keynesian Economics*, *4*(4), 373-390.

Bridji, S., & Charpe, M. (2015). *The impact of the labour share on growth: A time-frequency analysis* (Tech. Rep.). International Labour Organization.

Carvalho, L., & Rezai, A. (2016). Personal income inequality and aggregate demand. *Cambridge Journal of Economics*, *40*(2), 491-505.

Charpe, P., M.; Flaschel, & Proaño, C. (2016). *The persistence of goodwin profit/wage squeeze cycles in wage–led economies.* Mimeo, Bielefeld University, Germany.

Diks, C., & Panchenko, V. (2006). A new statistic and practical guidelines for non-parametric granger causality testing. *Journal of Economic Dynamics & Control*, *30*(9-10), 1647-1669.

Dutt, A. (1984). Stagnation, income–distribution and monopoly power. *Cambridge Journal of Economics*, *8*(1), 25-40.

Erturk, K. (2016). *Power hazard in market exchange: Can asymmetric power be harmful to all?* Mimeo, University of Utah, USA.

Feldstein, M. (2008). Did wages reflect growth in productivity? *Journal of Policy Modeling*, *30*(4), 591-594.

Flaschel, P. (2010). The classical growth cycle: Reformulation, simulation and some facts. In P. Flaschel (Ed.), *Topics in classical micro- and macroeconomics: Elements of a critique of neoricardian theory* (pp. 435–463). Berlin and Heidelberg: Springer.

Gallegati, M., Gallegati, M., Ramsey, J., & Semmler, W. (2011). The US wage phillips curve across frequencies and over time. *Oxford Bulletin of Economics and Statistics*, *73*(4), 489-508.

Gallegati, M., & Semmler, W. (2014). *Wavelet applications in economics and finance*. New York: Springer.

Goodwin, R. (1967). A growth cycle. In C. Feinstein (Ed.), *Socialism, capitalism and economic growth* (pp. 54–58). Cambridge, UK: Cambridge University Press.

Hein, E., & Vogel, L. (2008). Distribution and growth reconsidered: Empirical results for six OECD countries. *Cambridge Journal of Economics*, *32*(3), 479-511.

Hiemstra, C., & Jones, J. (1994). Testing for linear and nonlinear granger causality in the stock price–volume relation. *Journal of Finance*, *49*(5), 1639-1664.

Kiefer, D., & Rada, C. (2015). Profit maximising goes global: The race to the bottom. *Cambridge Journal of Economics*, *39*(5), 1333–1350.

Li, Y., & Shukur, G. (2011). Linear and nonlinear causality tests in an LSTAR model: Wavelet decomposition in a nonlinear environment. *Journal of Statistical Computation and Simulation*, *81*(12), 1913-1925.

Mohun, S., & Veneziani, R. (2008). Goodwin cycles and the US economy, 1948-2004. In P. Flaschel & M. Landesmann (Eds.), *Mathematical economics and the dynamics of capitalism: Goodwin's legacy continued* (pp. 107–130). London: Routledge.

Nikiforos, M. (2013). The (normal) rate of capacity utilization at the firm level. *Metroeconomica*, *64*(3), 513-538.

Nikiforos, M. (2016). On the 'utilisation controversy': A theoretical and empirical discussion of the kaleckian model of growth and distribution. *Cambridge Journal of Economics*, *40*(2), 437-467.

Onaran, O., & Galanis, G. (2012). *Is aggregate demand wage–led or profit–led? national and global effects.* (Tech. Rep.).

Percival, D. (1995). On estimation of the wavelet variance. *Biometrika*, *82*(3), 619-631.

Percival, D., & Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, *92*(439), 868-880.

Percival, D., & Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge ; New York: Cambridge University Press.

Piketty, T., & Saez, E. (2003). Income inequality in the united states, 1913-1998. *Quarterly Journal of Economics*, *118*(1), 1-39.

Pitarakis, J. (2006). Model selection uncertainty and detection of threshold effects. *Studies in Nonlinear Dynamics and Econometrics*, *10*(1).

Proaño, C., Flaschel, P., Ernst, E., & Semmler, W. (2009). Disequilibrium macroeconomic dynamics, income distribution and wage–price phillips curves: Evidence from the U.S. and the Euro Area. In P. Flaschel (Ed.), *The macrodynamics of capitalism: Elements for a synthesis of Marx, Keynes and Schumpeter* (Second revised and enlarged edition. ed., pp. xiv, 399 pages). Berlin ; London: Springer.

Rezai, A. (2015). Demand and distribution in integrated economies. *Cambridge Journal of Economics*, *39*(5), 1399-1414.

Rowthorn, R. (1982). Demand, real wwges and economic growth. *Studi Economici*, *18*, 3-53.

Sasaki, H., & Fujita, S. (2015). *Demand and income distribution in a two–country kaleckian model* (Tech. Rep.).

Skott, P. (2012). Theoretical and empirical shortcomings of the kaleckian investment function. *Metroeconomica*, *63*(1), 109-138.

Skott, P. (2017). Weaknesses of "wage–led growth". *Review of Keynesian Economics*, *5*(3), 336–359.

Skott, P., & Zipperer, B. (2012). An empirical evaluation of three post–keynesian models. *European Journal of Economics and Economic Policies: Intervention*, *9*, 277-308.

Tavani, D., Flaschel, P., & Taylor, L. (2011). Estimated non-linearities and multiple equilibria in a model of distributive-demand cycles. *International Review of Applied Economics*, *25*(5), 519–538.

Tavani, D., & Vasudevan, R. (2014). Capitalists, workers, and managers: Wage inequality and effective demand. *Structural Change and Economic Dynamics*, *30*, 120-131.

Taylor, L. (1985). A atagnationist model of economic–growth. *Cambridge Journal of Economics*, *9*(4), 383-403.

Taylor, L. (1991). *Income distribution, inflation, and growth : Lectures on structuralist macroeconomic theory*. Cambridge, Mass.: MIT Press.

von Arnim, R., & Barrales, J. (2015). Demand-driven Goodwin cycles with kaldorian and kaleckian features. *Review of Keynesian Economics*, *3*(3), 351–373.

Whitcher, B., Guttorp, P., & Percival, D. (1999). *Mathematical background for wavelet estimators of cross–covariance and cross–correlation* (Tech. Rep.). National Research Center for Statistics and the Environment.

Whitcher, B., Guttorp, P., & Percival, D. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research-Atmospheres*, *105*(D11), 14941-14962.

Zipperer, B., & Skott, P. (2011). Cyclical patterns of employment, utilization, and profitability. *Journal of Post Keynesian Economics*, *34*(1), 25-58.

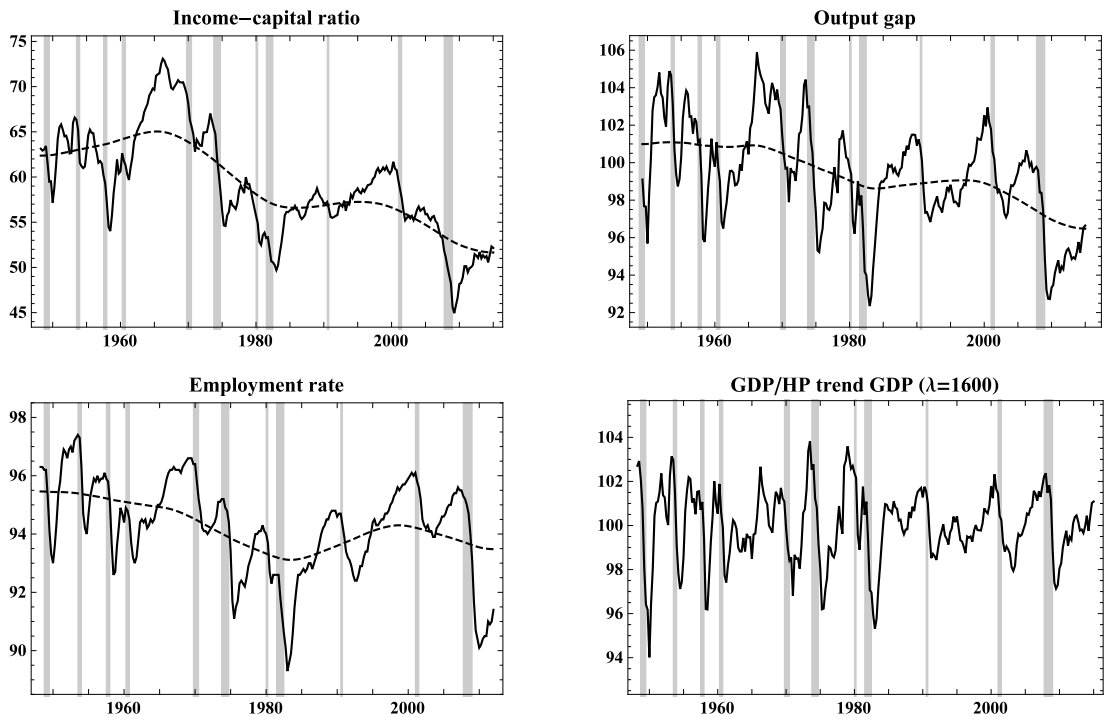


Figure 1.1: **Demand and employment**. Income–capital ratio, output gap, employment rate, and the ratio of real GDP to its Hodrick–Prescott trend. Shaded areas indicate NBER recession dates. See section 1.3 for further details on the series.

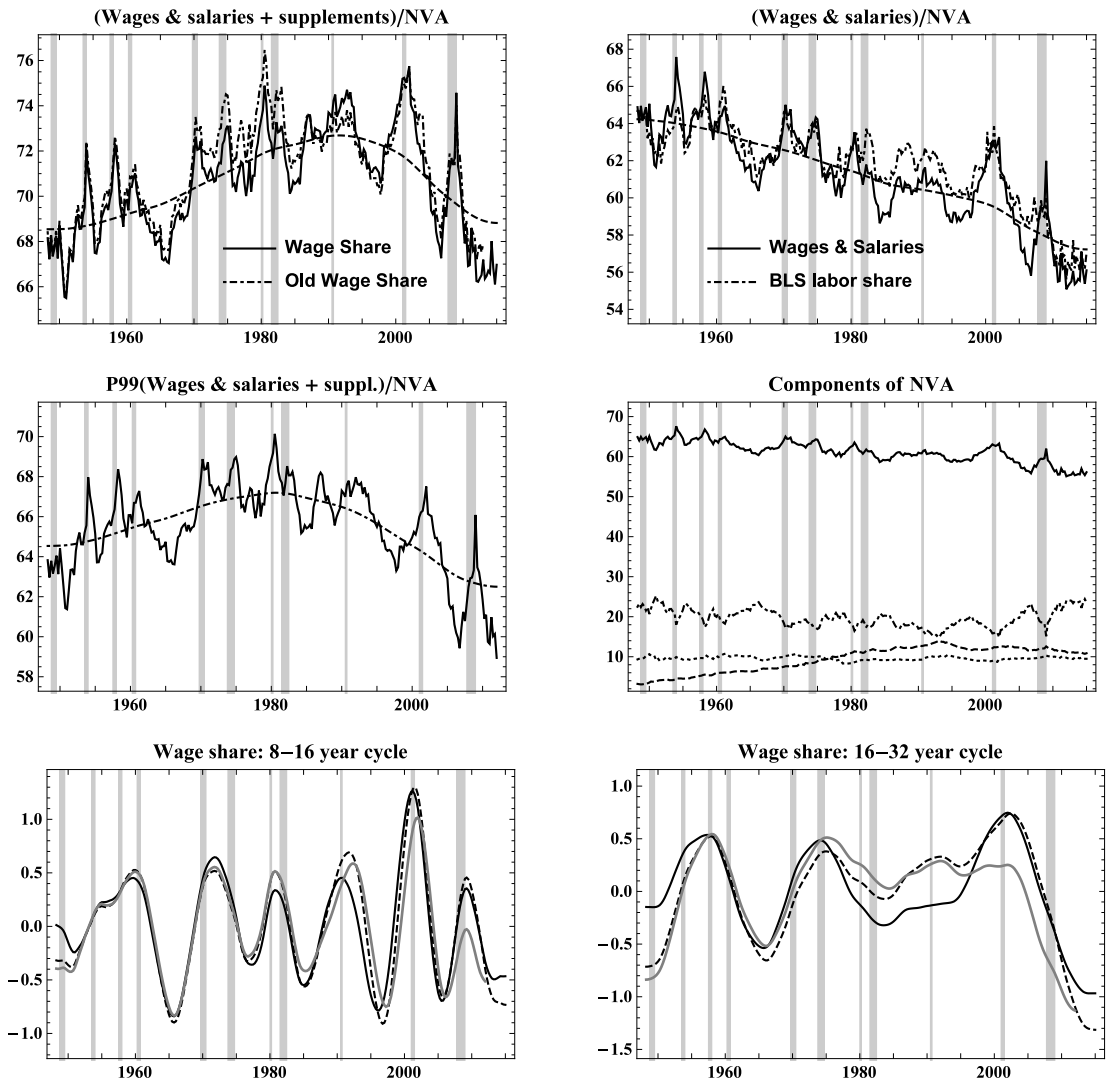


Figure 1.2: **Wage share**. Top row, left: compensation (wages and salaries plus supplements) of employees as a share of corporate net value added (NVA). Right: wages and salaries as a share of NVA. Middle row, left: "P99" is the wage share in the top left panel, adjusted by Piketty and Saez' share of the top 1 percent of wage income recipients. Right: components of NVA as shares of NVA; from top to bottom before 1960: wages and salaries, net operating surplus, taxes, supplements. Bottom row: wavelet decompositions of the three different wage share measures at two different periodicities (black: compensation/NVA; dashed: wages and salaries/NVA; gray: "P99" adjusted compensation/NVA). See section 1.3 for further discussion.

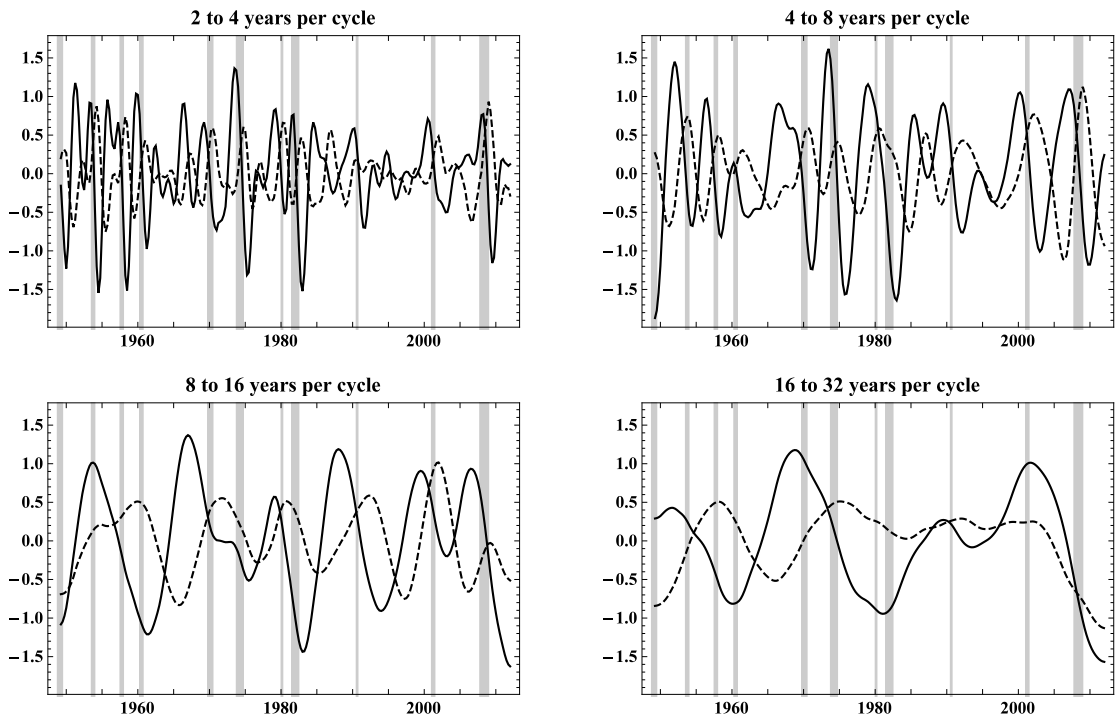


Figure 1.3: **Wavelet decomposition of output gap and wage share**. The dotted line represents the wage share (including supplements, minus top 1 percent), the solid line the output gap. Shaded areas indicate NBER recession dates. See section 1.3 for further details on the data, and the main text for details on the wavelet decomposition.

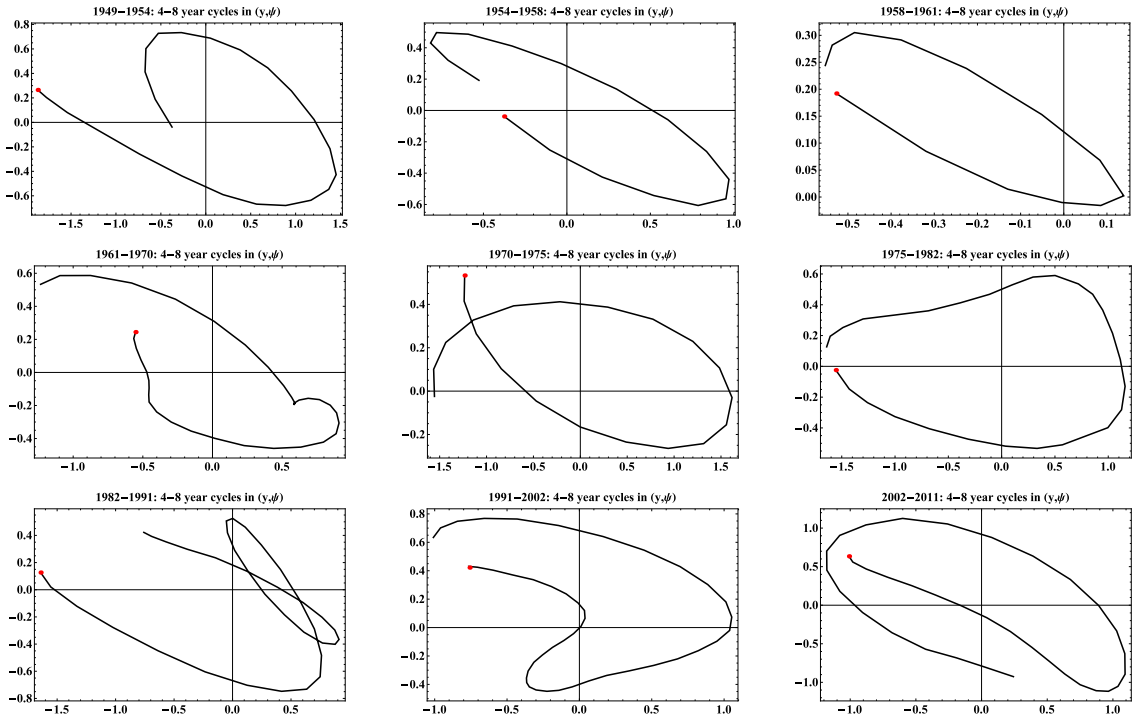


Figure 1.4: **Output gap and wage share, phase trajectories at business cycle frequency (4–8 years)**. The panels show wavelet decompositions for output gap (horizontal axis) and wage share (incl. supplements; minus top 1 percent). The dot indicates the starting point in each panel. See section 1.3 for further details on the series, and section 1.5 for discussion.

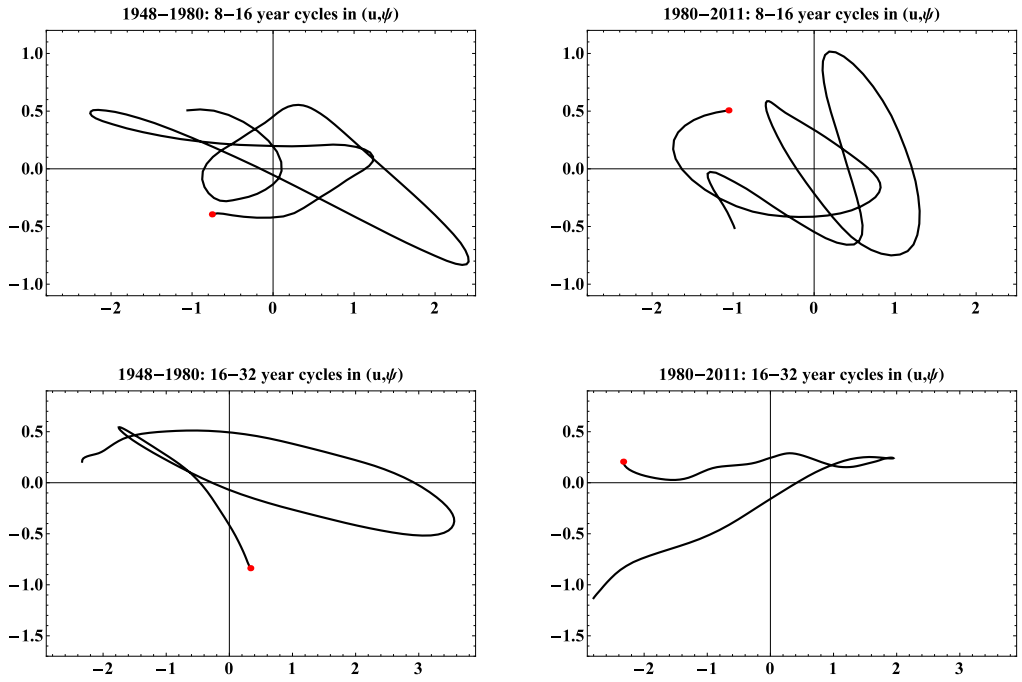


Figure 1.5: **Income–capital ratio and wage share, phase trajectories**. The panels show wavelet decompositions for income–capital ratio (horizontal axis) and wage share (incl. supplements; minus top 1 percent). The dot indicates the starting point in each panel. Top row displays 8–16 year cycles, pre– and post–1980; bottom row 16–32 year cycles. See section 1.3 for further details on the series, and section 1.5 for discussion.

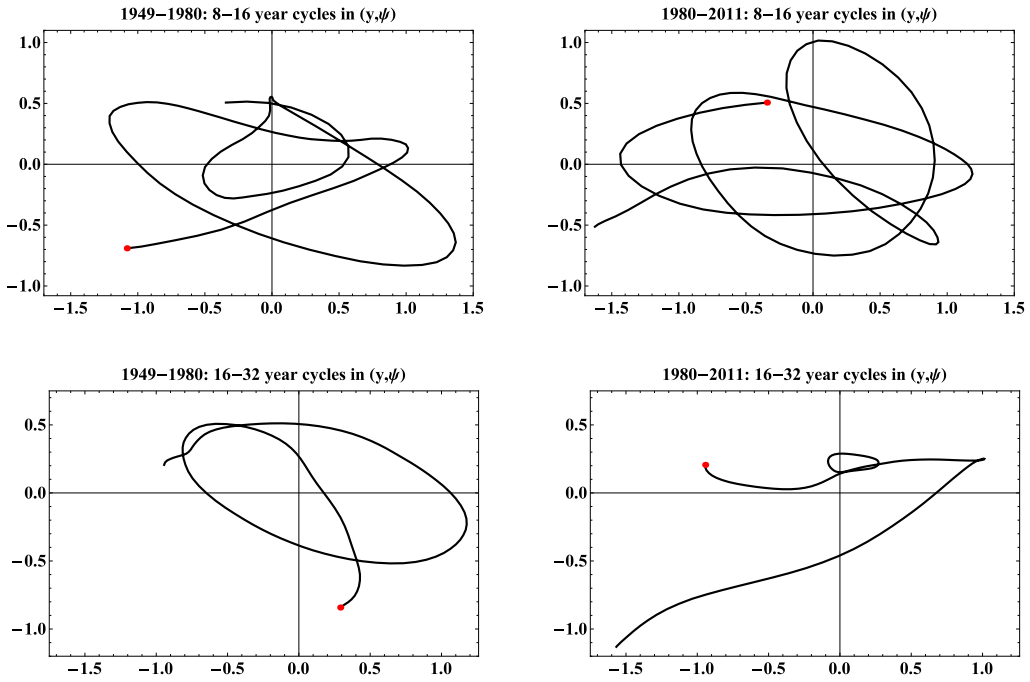


Figure 1.6: **Output gap and wage share, phase trajectories**. The panels show wavelet decompositions for output gap (horizontal axis) and wage share (incl. supplements; minus top 1 percent). The dot indicates the starting point in each panel. Top row displays 8–16 year cycles, pre– and post–1980; bottom row 16–32 year cycles. See section 1.3 for further details on the series, and section 1.5 for discussion.

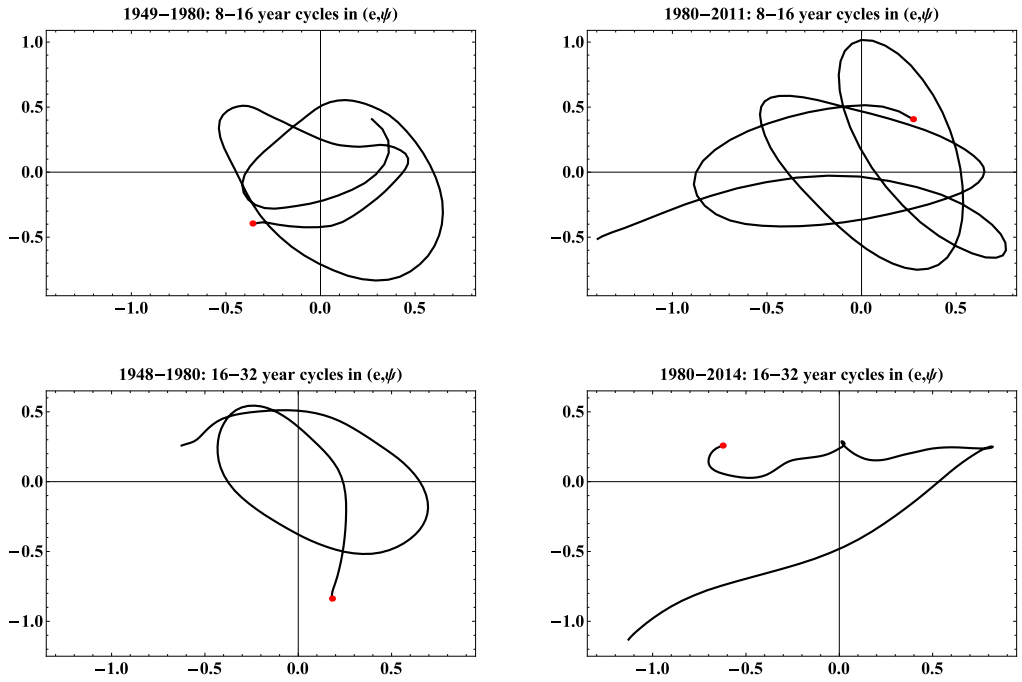


Figure 1.7: **Employment rate and wage share, phase trajectories**. The panels show wavelet decompositions for employment rate (horizontal axis) and wage share (incl. supplements; minus top 1 percent). The dot indicates the starting point in each panel. Top row displays 8–16 year cycles, pre– and post–1980; bottom row 16–32 year cycles. See section 1.3 for further details on the series, and section 1.5 for discussion.

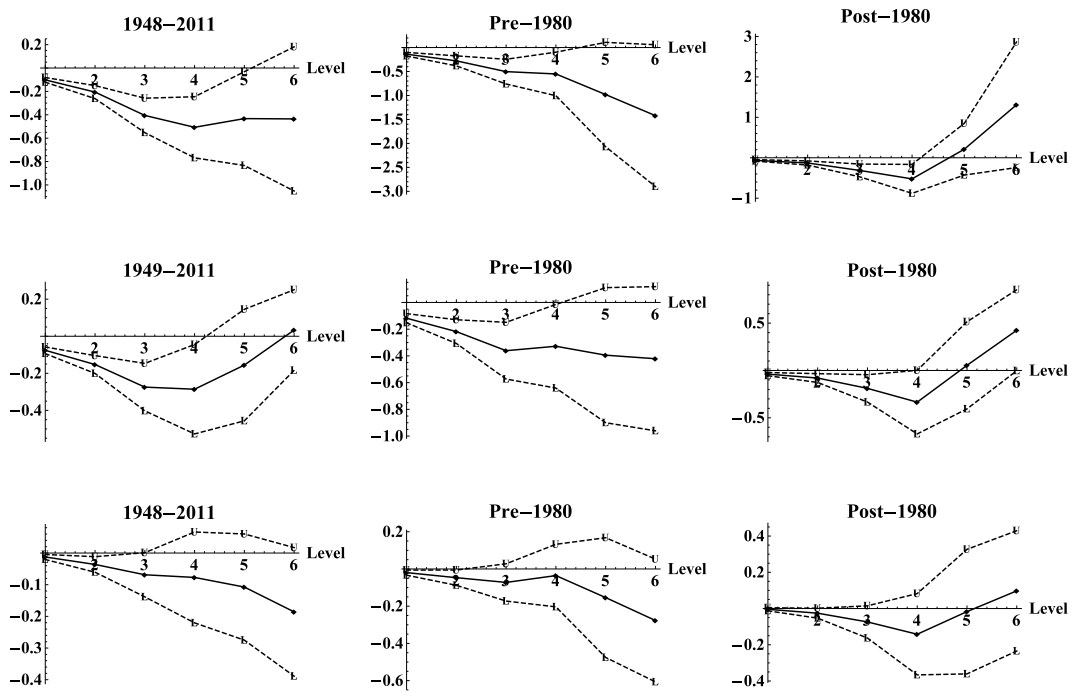


Figure 1.8: **Wavelet covariances**. Top row: covariance of income–capital ratio and wage share at time scales 1–6. Middle row: covariance of output gap and wage share at time scales 1–6; bottom row: covariance of employment rate and wage share at time scales 1–6. The left-most panels show covariances for the entire sample period, the middle panels for the years before 1980, and the right panels for the period 1981–2014.

Table 1.1: **Nonlinear Granger causality test** ($\varepsilon = 1.5$).

			Income–capital		**Output gap**		**Employment rate**	
		lags	T–test	p–value	T–test	p–value	T–test	p–value
0.5–1	$\psi \Rightarrow ...$	2	0.6021	0.2736	-0.8661	0.1932	1.2995^{*}	0.0969
	$... \Rightarrow \psi$		0.3177	0.3754	2.1741^{**}	0.0149	1.8070^{**}	0.0354
1–2	$\psi \Rightarrow ...$	4	2.3254^{**}	0.0100	2.4529^{***}	0.0071	1.3612^{*}	0.0867
	$... \Rightarrow \psi$		1.6571^{**}	0.0487	2.0118^{**}	0.0221	2.5533^{***}	0.0053
2–4	$\psi \Rightarrow ...$	8	2.6893^{***}	0.0036	2.9942^{***}	0.0014	2.8329^{***}	0.0023
	$... \Rightarrow \psi$		2.0096^{**}	0.0222	2.4037^{***}	0.0081	2.0838^{**}	0.0186
4–8	$\psi \Rightarrow ...$	16	2.0473^{**}	0.0203	2.4728^{***}	0.0067	2.5002^{***}	0.0062
	$... \Rightarrow \psi$		1.1814	0.1187	1.9602^{**}	0.0250	1.8897^{**}	0.0294
8–16	$\psi \Rightarrow ...$	32	2.3363^{***}	0.0097	3.2873^{***}	0.0005	3.3833^{***}	0.0004
	$... \Rightarrow \psi$		1.9216^{**}	0.0273	1.1812	0.1188	1.1708	0.1208
16–32	$\psi \Rightarrow ...$	64	1.3139^{*}	0.0944	2.6752^{***}	0.0037	1.8296^{**}	0.0337
	$... \Rightarrow \psi$		1.4191^{*}	0.0779	1.3671^{*}	0.0858	1.5716^{*}	0.0580

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

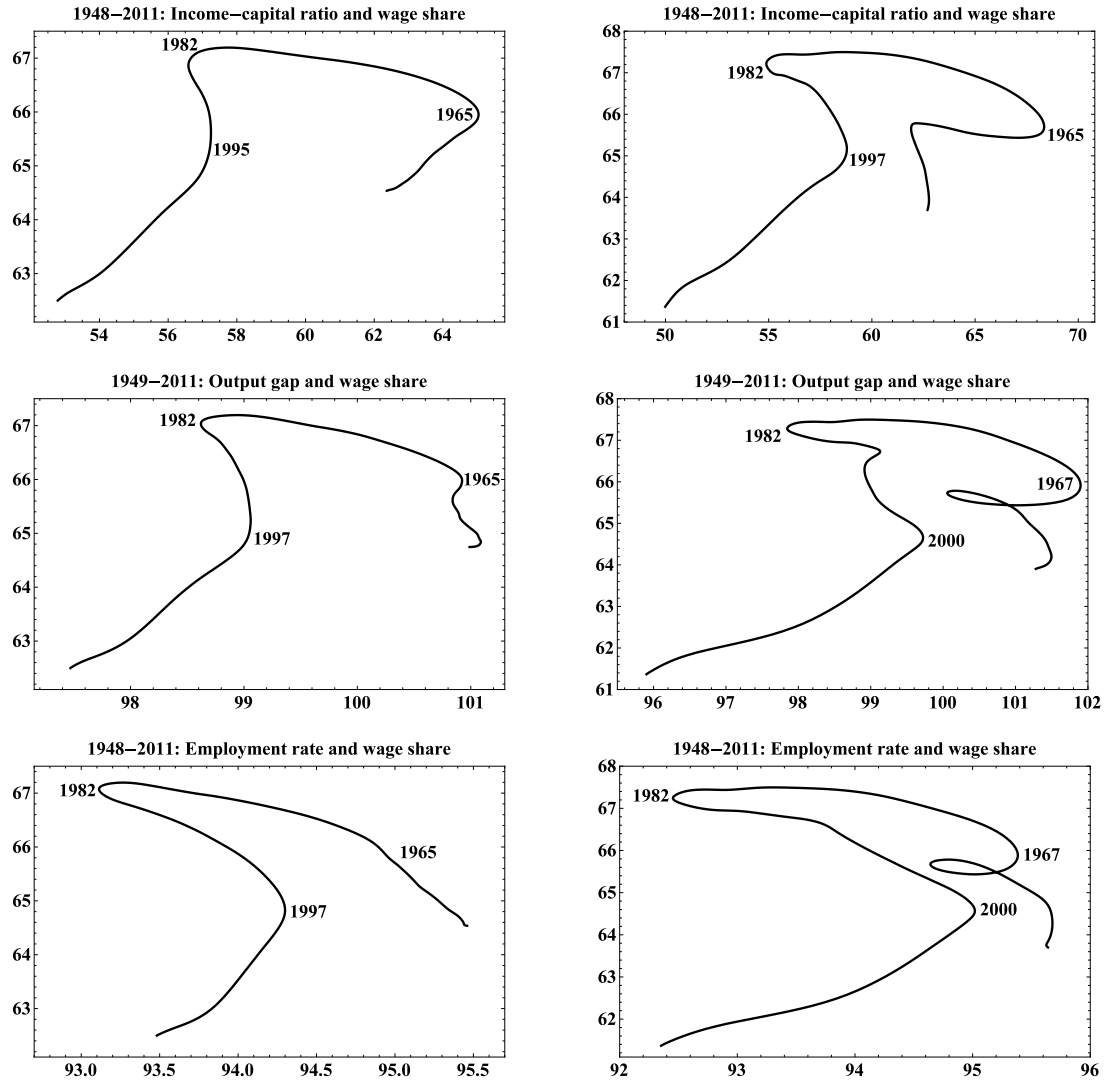


Figure 1.9: **Income–capital ratio, output gap, and employment rate vis–à–vis the wage share, trends 1948/49–2011**. In all panels, the wage share (incl. supplement; minus top 1 percent) is on the vertical axis. Panels on the left are based on "smooth trends" with periodicity greater than 128 quarters per cycle (or 32 years per cycle); panels on the right include trends with 64 quarters per cycle. See section 1.3 for further details on the series, and section 1.5 for discussion.

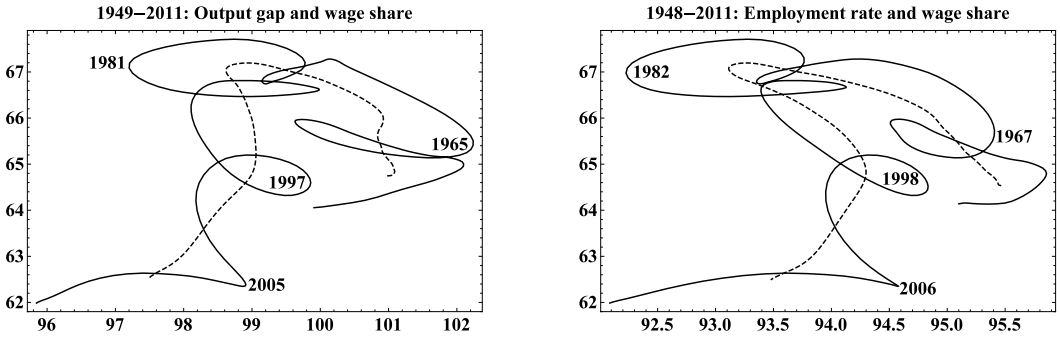


Figure 1.10: **Output gap and employment rate vis–à–vis the wage share, trends and cycles 1948/49–2011**. In both panels, the wage share (incl. supplement; minus top 1 percent) is on the vertical axis. The dashed lines are smooth trends as in Figures 1.1 and 1.2 (32+ years per cycle); the solid line represents cycles with 4–8 year cycles. The solid output gap–wage share cycles in the left panel are as well shown in Figure 1.4. See section 1.3 for further details on the series, and section 1.5 for discussion.

CHAPTER 2

ON THE SEARCH FOR ENDOGENOUS CYCLES: THE CASE OF WAGE SHARE AND OUTPUT GAP

2.1 Abstract

This paper seeks to investigate the individual dynamics of two important variables in the post–Keynesian literature that represent the functional income distribution and economic activity, i.e., wage share and output gap. Our results suggest that output gap behaves quasi chaotically, i.e., dominant Lyapunov exponent (LE) close to zero, and wage share noisy stable, i.e., negative LE.

2.2 Introduction

Post–Keynesian macroeconomic theories of business cycles are rich in endogenous mechanisms that lead to sustained oscillations. Seminal contributions are, for instance, Kalecki (1937), Kaldor (1940), and Goodwin (1951, 1967). These papers, in general, depict the economic system as having internal mechanisms that create cycles even in the absence of external shocks. This contrasts with the mainstream Real Business Cycles (RBC) models (Kydland & Prescott, 1982), and its modern incarnation Dynamic Stochastic General Equilibrium (DSGE) models (Christiano, Eichenbaum, & Evans, 2005; Smets & Wouters, 2007), which rely on the assumption that the capitalist economy is stable and economic fluctuations arise from stochastic shocks[1].

In search of endogenous mechanism, chaos theory seems appealing (e.g., Chen, 2010). From a theoretical perspective, first, chaotic behavior can be defined as sensitivity to initial conditions, implying that there is a limit on the predictability of future events, even in the

[1]Chen (2010, chapter 12) provides a history of the rise and success of the Frisch (1933) research agenda on noise–driven damped oscillations. He also provides a discussion in how the first–differencing filter method has been the main support of the exogenous cycle tradition, (e.g., Engle & Granger, 1987; Nelson & Plosser, 1982). First–differenced data amplifies noise and whitened the data, hence producing the illusion of equilibrium. Although to the naked eye the difference between a chaotic and noise driven time series cannot be distinguished, their power spectra are completely different.

absence of noise; second, the stretching and folding mechanisms, which are at the core of chaos, imply strong internal dynamics that sustain the oscillations; and third, complex phenomena might be described by simple rules or by low dimensional dynamics. From an empirical standpoint, traditional regression and spectral techniques, although able to study some features of the data and predict it, are not able to reveal its deterministic underlying structure (Wolff, 1992).

An operational definition of chaos is *systems that have bounded solutions and dominant Lyapunov exponent (λ) larger than zero*, i.e., two slightly different initial conditions depart from each other exponentially. Usually, this definition applies to dynamical systems not affected by noise, and therefore, some authors claim that there is a striking difference between stochastic and deterministic systems (see, for example, Dennis, Desharnais, Cushing, Henson, & Costantino, 2003). However, in real systems, endogenous feedback mechanisms and noise go hand in hand (Coulson, Rohani, & Pascual, 2004). Therefore, the relevant question seems to be: *Do endogenous feedback mechanisms contribute to the observed irregularity, and if so, to what degree?* Ellner and Turchin (2005) proposed a four–fold category regarding the relationship between noise (R^2) and the dominant Lyapunov exponent (λ): noisy chaos, noisy stability, quasi–chaos, and noise–dominated situations.

Noisy chaos, $\lambda > 0$ and high R^2, represents a system that amplifies the noise by its internal feedback mechanisms. Noisy stability, $\lambda < 0$ and high R^2, stands for a stable system that damps the noise. Quasi chaos, $\lambda \approx 0$ and high R^2, is a system that moves between stable and chaotic behavior. A noisy system can jump from stable to chaotic regime smoothly. This is not true in the absence of noise. In these circumstances, the distribution of the so–called finite time or local Lyapunov exponent (LLE) becomes relevant (see subsection 2.3.1 for more detail). Finally, noise dominated dynamics, where a low R^2 makes that the dominant Lyapunov exponent plays no relevant role, because the internal dynamics are irrelevant.

Following this classification, output gap is found to behave quasi–chaotically; meanwhile, wage share is noisy stable.

The remainder of the paper is structure as follows: section 2.3 describes the dominant and local Lyapunov exponent, as well as the univariate Nonlinear Autoregressive Neural Network (NARNN) model used to calculate them. In section 2.4, We use the above meth-

ods to study wage share and output gap independently. Concluding remarks are in section 2.5.

2.3 Lyapunov exponents

One main feature of a chaotic system is that starting from slightly different conditions, the evolution of the paths departs from each other exponentially. The quantity that measures that is the dominant Lyapunov exponent. The usual classification is a fixed point if the exponents are negative, limit cycle if one exponent is zero, and the remaining are negative, k–dimensional torus (quasi–periodic motion) if k exponents are zero and all the remaining are negative and strange attractor if at least one is positive.

Specifically, consider the discrete differentiable dynamical system subject to noise[2].

$$X_{t+1} = M(X_t) + E_{t+1}, \tag{2.1}$$

where $X_t, E_t \in \mathbb{R}^p$ for integers t and M is a $\mathbb{R}^p \to \mathbb{R}^p$ map. E_t are iid with zero mean, covariance Σ, and independent from X_0.

Let $J(X)$ be Jacobian matrix of M at X. The separation between two close initial values X_0 and X_0' after time n is given approximately by

$$X_n - X_n' = M^n(X_0) - M^n(X_0') \approx J^n(X_0)(X_0 - X_0'), \tag{2.2}$$

By the chain rule of differentiation,

$$J^n(X_0) = J(X_{n-1})J(X_{n-2})\dots J(X_0). \tag{2.3}$$

Let $||\bullet||$ denote the matrix norm, then the Lyapunov exponents are given by

$$\lambda_n = \lim_{n\to\infty} n^{-1} \ln ||J^n(X_0)||, \tag{2.4}$$

There are two main methods to calculate Lyapunov exponents: direct and indirect (Jacobian methods). The former compares the evolution of two similar initial points. This method is almost forbidden for a typical economic time series because it requires a large amount of observations and assumes that they are noiseless, or at least that they are subject to the *same* noise. Using this method in small and noisy data sets could lead to wrong conclusions (McCaffrey, Ellner, Gallant, & Nychka, 1992).

[2]The dynamical system (2.1) is called differentiable if the map M is differentiable.

The indirect method, on the other hand, consists of fitting a nonparametric regression to the data to calculate the Jacobian matrices and estimate the LE by the QR–factorization described in Appendix B.4. One important difficulty lies here: the dimension of the attractor is required, so we need all the interacting variables in the model. However, we are never certain of which variable should be included to form the attractor. Fortunately, it is possible to use time–delay coordinates, known as *attractor reconstruction* (Sauer, Yorke, & Casdagli, 1991; Takens, 1981), that have the same properties as the original attractor. If the true attractor lies on an p-dimensional space, then for almost every *embedding lag* τ and large enough *embedding dimension* m, the attractor of the m-dimensional time series

$$X_t = \{x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(m-1)\tau}\} \tag{2.5}$$

is qualitatively similar to the unknown attractor of the p-dimensional system (Pascual & Ellner, 2000). The embedding dimension (m) must be, at least, larger than p.

2.3.1 Local Lyapunov exponents

The dominant Lyapunov exponent can be considered as the long run *average* of divergence between two close points affected by a small perturbation. Furthermore, by Oseledets' multiplicative ergodic theorem, it is independent from the initial conditions. Hence, it studies the asymptotic behavior of the system. However, transient behavior, which cannot be captured by asymptotic tools, might play a crucial role in the dynamics.

This turns out to be crucial when $\lambda \approx 0$ because in principle, any dynamics might be possible[3]. For that reason, we use the so–called finite time or local Lyapunov exponent (LLE). LLE provides information on how a perturbation to a system's orbits will *exponentially* increase or decrease in *finite time*, therefore indicating the predictability of a system in the short run (Bailey, 1996; Turchin & Ellner, 2000; Wolff, 1992; Ziehmann, Smith, & Kurths, 2000). LLE can be written as

$$\lambda(\omega, X_0) = \omega^{-1} \ln ||J^{\omega}(X_0)||, \tag{2.6}$$

and is calculated as well as LE using the QR–factorization (see Appendix B.4).

Equation (2.6) makes explicit the fact that, contrary to the dominant Lyapunov exponent, LLE depends on the initial conditions (X_0) and the time window (ω). Therefore,

[3] $\lambda \approx 0$ might be regarded as λ between -0.1 and 0.1.

its probability distribution depends on these two parameters (see Prasad & Ramaswamy, 1999). If ω is chosen too small, the distribution will keep changing shape, because it has not reached its stationary distribution. If chosen to be too long, the distribution will be a δ function at the dominant Lyapunov exponent. For the present study, we choose $\omega = 25$ quarters.

2.3.2 NARNN: Estimating the Jacobian

In order to estimate the Jacobian matrix, $J(X_t)$, we rewrite the model (2.1), with the reconstructed attractor (see equation (2.5), with $\tau = 1$), in its state space form,

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-m+1} \end{pmatrix} = \begin{pmatrix} F(X_{t-1}) \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-m+1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.7}$$

with F an unknown function that has to be approximated. The Jacobian matrix is given by,

$$J(X_t) = \begin{pmatrix} \Delta F_{x_{t-1}} & \Delta F_{x_{t-2}} & \dots & \Delta F_{x_{t-m+1}} & \Delta F_{x_{t-m}} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \tag{2.8}$$

where ΔF_{x_j} is the partial derivative of F with respect to x_j.

The estimation is carried out using the Nonlinear AutoRegressive Neural Network (NARNN) model in Matlab's Neural Network Toolbox by the Levenberg-Marquardt back-propagation algorithm described in Appendix B.5. NARNN's series–parallel method (or open loop) approximates the map $F(\bullet)$ by a feedforward neural network with single output and delayed outputs. Artificial Neural Networks (ANN) have the universal approximation property, i.e., they are able to approximate any function to any degree of accuracy and *its derivatives* (Gallant & White, 1992; Hornik, Stinchcombe, & White, 1989). Furthermore, using NARNN to calculate the Jacobian matrix makes explicit that the model is believed to be subject to *dynamical* shocks. Therefore, there is interplay between the interaction of the deterministic skeleton and how forcing shocks could impact the system. This noise can play an important role defining changes in regime in the dynamics of a time series (Ellner & Turchin, 2005). It can cause anything from a mild blur on the attractor to

more serious effects (Kendall, 2001). For instance, if the dynamical system has multiple equilibria, shocks might cause jumps among them.

An example of a NARNN architecture is in Figure 2.1 (see Beninca et al., 2008, supplementary information). Generally, ANN have three layers: input, hidden, and output layers. The input layer consists, in this case, of the lagged values of the time series (neurons). Each neuron is then passed and combined in a hidden layer. Between the input layer and the hidden layer, a linear transformation is performed, i.e.,

$$\text{hidden}_j = \sum_{i=1}^{m} x(t-i)\gamma_{ij} + \mu_j, \tag{2.9}$$

where $x(t)$ are input neurons, $\{\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{mj}\}$ the connection strength, and μ_j the intrinsic activity level. The hidden layer executes a nonlinear transformation,

$$\varphi(\text{hidden}_j) = \tanh(\text{hidden}_j), \tag{2.10}$$

which is a sigmoid function. It is an *activation function*, and bounded by -1 and 1. All the activation functions are collected in one linear transformation to calculate the output, with the form

$$F(X) = \beta_0 + \sum_{j=1}^{K} \beta_j \varphi\left(\text{hidden}_j\right). \tag{2.11}$$

The complexity of the model is chosen based on generalized cross–validation, GCV(α), i.e.,

$$\text{GCV}(\alpha) = \left(\frac{\text{RMSE}}{1 - \alpha\frac{k}{T}}\right)^2, \tag{2.12}$$

where RMSE is the root mean square error, T is the number of data points, and k is the number of parameters. We use GCV because it provides more flexibility than AIC and BIC. For my specific sample size and number of parameters, BIC tends to underfit and AIC to overfit; therefore, a good balance solution is to use GCV with $\alpha = 1.4$. After choosing a model based on GCV(1.4), the adequacy of the model is studied with the test developed by Billings and Zhu (1995) that tests for the unpredictability of residuals from, in our case, all linear and nonlinear combinations of past outputs and residuals. For our purposes, it can be written as:

$$\phi_{\varepsilon^2, x\varepsilon}(\tau) = \lambda_0 \delta(\tau) \tag{2.13}$$

where δ the delta function, $\lambda_0 = \sqrt{\sum_{t=1}^{T}(\varepsilon(t)^2 - \overline{\varepsilon(t)^2})} \Big/ \sqrt{\sum_{t=1}^{T}(x(t)\varepsilon(t) - \overline{x(t)\varepsilon(t)})}$ a constant, $\overline{x}$ the average of x, and ϕ_{xy} the cross–correlation function between x and y, i.e.,

$$\phi_{xy}(\tau) = \frac{\sum_{t=1}^{N-\tau}[x(t) - \overline{x}]\,[y(t+\tau) - \overline{y}]}{\sqrt{\sum_{t=1}^{N}[x(t) - \overline{x}]^2}\sqrt{\sum_{t=1}^{N}[y(t) - \overline{y}]^2}}. \tag{2.14}$$

with confidence interval at 95% between $\pm 1.96/\sqrt{T}$.

2.4 Data analysis

For the functional distribution of income and wage share, we use the corporate labor cost as a share of corporate net value added from BEA NIPA table 1.14, from 1948:I to 2014:IV. A detailed discussion can be found in Chapter 1. The economic activity is measure by output gap. Figure 2.2 upper panel shows, left, the log of wage share and, right, the log of the real output and their respective trends. The bottom panel exhibits the detrended wage share (ψ) and the output gap (u) both calculated as the percentage deviation from their trend. The trend is computed using the Maximal–Overlap Discrete Wavelet Transform (MODWT) for periods longer than 64 quarters per cycle (see Appendix B.1 and Percival and Walden (2000) for more detail).

Trend is measured as periodicities longer than 64 quarters per cycle. Although this definition seems unusual, it has been shown by Comin and Gertler (2006) and Pancrazi (2015) that medium–run dynamics larger than 32 quarters per cycle, usually considered as a trend, greatly influence business cycles fluctuations. The trend is then calculated with the Maximal Overlap Discrete Wavelet Transform (MODWT). In a nutshell, MODWT for level J for a time series $\boldsymbol{X}$ yields highly redundant and non–orthogonal column vectors $\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \cdots, \tilde{\boldsymbol{W}}_J$ and $\tilde{\boldsymbol{V}}_J$ each of dimension T. Vectors $\tilde{\boldsymbol{W}}_j$ are associated with an approximated bandpass filter with frequency $[2^{-(j+1)}, 2^{-j}]$ (or periodicities between $[2^j, 2^{j+1}]$ quarters per cycle), meanwhile $\tilde{\boldsymbol{V}}_J$ is associated with a lowpass filter with approximate frequencies $[0, 2^{-(J+1)}]$ (or periodicities larger than 2^{J+1} quarters per cycles).

MODWT yields an energy decomposition

$$||\boldsymbol{X}||^2 = \sum_{j=1}^{J} ||\tilde{\boldsymbol{W}}_j||^2 + ||\tilde{\boldsymbol{V}}_J||^2, \tag{2.15}$$

where $||\bullet||$ is the l^2–norm, and an additive decomposition called MultiResolution Analysis (MRA)

$$\boldsymbol{X} = \sum_{j=1}^{J} \tilde{\mathfrak{D}}_j + \tilde{\mathfrak{S}}_J, \tag{2.16}$$

where $\tilde{\mathfrak{D}}_j$ and $\tilde{\mathfrak{S}}_J$ are the j–th order *detail* and the J–th order *smooth* for $\boldsymbol{X}$, respectively. The transform from $\boldsymbol{X}$ to $\tilde{\boldsymbol{W}}_j$ and from $\boldsymbol{X}$ to $\tilde{\boldsymbol{V}}_j$ can be expressed as

$$\tilde{\boldsymbol{W}}_j = \tilde{\mathcal{W}}_j \boldsymbol{X} \quad \text{and} \quad \tilde{\boldsymbol{V}}_j = \tilde{\mathcal{V}}_j \boldsymbol{X}, \tag{2.17}$$

where each row of the $N \times N$ matrix $\tilde{W}_j$ has a value of the periodized filter to length N of $\{\tilde{h}_j\}$ and $\{\tilde{g}_j\}$, which are called wavelet and scaling filters, respectively. The MRA is obtained, therefore, as

$$\tilde{\mathfrak{D}}_j \equiv \tilde{\mathcal{W}}_j^T \tilde{\boldsymbol{W}}_j \quad \text{and} \quad \tilde{\mathfrak{S}}_j \equiv \tilde{\mathcal{V}}_j^T \tilde{\boldsymbol{V}}_j, \tag{2.18}$$

where $\tilde{\mathfrak{S}}_J$ is defined as the trend of $\boldsymbol{X}$. We choose the Daubichies' least asymmetric (LA) wavelet filter LA(8) and $J = 5$.

A first indicator to assess the possibility of chaos in the data is given by the distribution of their power spectrum because a broad band spectrum is a necessary, but not sufficient, condition for chaos. For this task, we use the continuous wavelet transform (CWT). CWT maps the time series from time to timescale domain, providing time–frequency *localized* information. This allows the study of the evolution of the power spectrum and some transient state information that otherwise would be lost using, for instance, Fourier analysis (for details see Appendix B.2). This is clear in Figure 2.3, which shows the wavelet power spectrum (WPS) and the global wavelet power spectrum (GWPS). The GWPS is the time average of the WPS (see Appendix B.2, equation (B.18)) which resembles the Fourier spectral density. It clearly shows that both wage share and capacity utilization are describe by a broad band power spectrum. On the other hand, WPS provides timescale localized information of the distribution of the variance: warmer colors represent higher power, i.e., red the highest and blue the lowest. White lines show the highest level of power at each time period, the so–called *characteristic period*.

2.4.1 Estimation results

The sample considered for the neural network estimation is 1956:I-2004:IV to avoid both the 2008's Great Recession and boundary effect from MODWT filtering. Each time series is also normalize between -1 and 1 and divided 80% for training and 20% for validation. We start the discussion of the results with the wage share. Table 2.1 presents the

neural network results. Residuals are not correlated with past outputs and past residuals (Figure 2.4 left panel). The R^2 is 88.70% and LE is -0.3607 with confidence interval -0.6010 and -0.1820 at 10%. Figure 2.5 presents the bootstrap distribution of λ (see Appendix B.3 for details on bootstrapping). Therefore, the system seems noisy–stable. Figure 2.6 left panel displays the evolution of 25–step–ahead LLE, which reveals how wage share becomes more predictable when reaching a peak, i.e., in any major peak, LLE becomes more negative.

Regarding output gap, Table 2.1 summarizes the estimation results. Residuals are white noise (Figure 2.4 right panel). The R^2 is 93.93% and LE is -0.0657, with confidence interval between -0.1250 and -0.0067 at 10% of significance. Based on these results, output gap seems to behave quasi–chaotically: see Figure 2.5. In the short run, see Figure 2.6 right panel, the 25–step–ahead LLE shows that both major peaks and troughs are more predictable.

Moreover, Appendix A presents the same analysis for two additional measurement of economic activity, i.e., income–capital ratio and GDP to potential GDP calculated by the Congressional Budget Office (CBO), both defined in detail in Chapter 1 and Appendix A. Both measurements seem to follow a quasi–chaotic behavior since LE is close to zero.

2.5 Concluding remarks

In this paper, we study the empirical dynamical properties of a measurement of functional income distribution and economic activity, i.e., wage share and output gap. My specific objective is to obtain further insights regarding the possible nonlinear behavior of these variables. For that reason, we use the NARNN model to estimate, by means of Takens' embedding theorem, the dominant and local Lyapunov exponent of the time series. The results indicate that while wage share follows a noisy–stable dynamics, i.e., negative LE (less than -0.1), output gap follows a quasi-chaotic behavior. Furthermore, the investigation of the LLE reveals that during peaks, wage share and output gap become more predictable.

2.6 References

Bailey, B. (1996). Local Lyapunov exponents: Predictability depends on where you are. In W. A. Barnett, A. P. Kirman, & M. Salmon (Eds.), *Nonlinear dynamics and economics*

(pp. 345–359). Cambridge ; New York: Cambridge University Press.

Beninca, E., Huisman, J., Heerkloss, R., Johnk, K., Branco, P., Van Nes, E., ... Ellner, S. (2008). Chaos in a long–term experiment with a plankton community. *Nature*, *451*(7180), 822–825.

Billings, S., & Zhu, Q. (1995). Model validation tests for multivariable nonlinear models including neural networks. *International Journal of Control*, *62*(4), 749-766.

Chen, P. (2010). *Economic complexity and equilibrium illusion: Essays on market instability and macro vitality* (1st ed.). New York and London: Routledge.

Christiano, L., Eichenbaum, M., & Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, *113*(1), 1–45.

Comin, D., & Gertler, M. (2006). Medium–term business cycles. *American Economic Review*, *96*(3), 523–551.

Coulson, T., Rohani, P., & Pascual, M. (2004). Skeletons, noise and population growth: the end of an old debate? *Trends in Ecology & Evolution*, *19*(7), 359–364.

Dennis, B., Desharnais, R., Cushing, J., Henson, S. M., & Costantino, R. (2003). Can noise induce chaos? *Oikos*, *102*(2), 329–339.

Ellner, S. P., & Turchin, P. (2005). When can noise induce chaos and why does it matter: a critique. *Oikos*, *111*(3), 620–631.

Engle, R., & Granger, C. (1987). Co–integration and error correction: Representation, estimation, and testing. *Econometrica*, *55*(2), 251–276.

Foresee, F. D., & Hagan, M. T. (1997, Jun). Gauss–newton approximation to bayesian learning. In *Proceedings of the IEEE international conference on neural networks.* doi: 10.1109/ICNN.1997.614194

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics. In *Economic essays in honour of Gustav Cassel* (pp. 171–206). London: Allen & Unwin Ltd.

Gallant, A., & White, H. (1992). On learning the derivatives of an unknown mapping with multilayer feedforward networks. *Neural Networks*, *5*(1), 129–138.

Goodwin, R. (1951). The nonlinear accelerator and the persistence of business cycles. *Econometrica*, *19*(1), 1–17.

Goodwin, R. (1967). A growth cycle. In C. Feinstein (Ed.), *Socialism, capitalism and economic growth* (pp. 54–58). Cambridge, UK: Cambridge University Press.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Kaldor, N. (1940). A model of the trade cycle. *The Economic Journal*, *50*(197), 78–92.

Kalecki, M. (1937). A theory of the business cycle. *The Review of Economic Studies*, *4*(2), 77–97.

Kendall, B. (2001). Cycles, chaos, and noise in predator–prey dynamics. *Chaos*, *12*(2), 321–332.

Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, *50*(6), 1345–1370.

MacKay, D. (1992). Bayesian interpolation. In C. R. Smith, G. J. Erickson, & P. O. Neudorfer (Eds.), *Maximum entropy and bayesian methods: Seattle, 1991* (pp. 39–66). Dordrecht: Springer Netherlands.

McCaffrey, D., Ellner, S., Gallant, A., & Nychka, D. (1992). Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. *Journal of the American Statistical Association*, *87*(419), 682–695.

Nelson, C., & Plosser, C. (1982). Trends and random walks in macroeconmic time series. Some evidence and implications. *Journal of Monetary Economics*, *10*(2), 139–162.

Pancrazi, R. (2015). The heterogeneous great moderation. *European Economic Review*, *74*, 207–228.

Pascual, M., & Ellner, S. (2000). Linking ecological patterns to environmental forcing via nonlinear time series models. *Ecology*, *81*(10), 2767–2780.

Percival, D., & Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, *92*(439), 868-880.

Percival, D., & Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge ; New York: Cambridge University Press.

Prasad, A., & Ramaswamy, R. (1999). Characteristic distributions of finite–time Lyapunov exponents. *Physical Review E*, *60*(3), 2761–2766.

Sauer, T., Yorke, J., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, *65*(3), 579–616.

Smets, F., & Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, *97*(3), 586–606.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, Warwick 1980: Proceedings of a symposium held at the University of Warwick 1979/80* (pp. 366–381). Berlin, Heidelberg: Springer Berlin Heidelberg.

Turchin, P., & Ellner, S. (2000). Living on the edge of chaos: Population dynamics of Fennoscandian Voles. *Ecology*, *81*(11), 3099–3116.

von Bremen, H. F., Udwadia, F. E., & Proskurowski, W. (1997). An efficient QR based method for the computation of lyapunov exponents. *Physica D*, *101*(1), 1–16.

Wolff, R. (1992). Local Lyapunov exponents: Looking closely at chaos. *Journal of the Royal Statistical Society. Series B (Methodological)*, *54*(2), 353–371.

Ziehmann, C., Smith, L., & Kurths, J. (2000). Localized Lyapunov exponents and the prediction of predictability. *Physics Letters, Section A: General, Atomic and Solid State Physics*, *271*(4), 237–251.

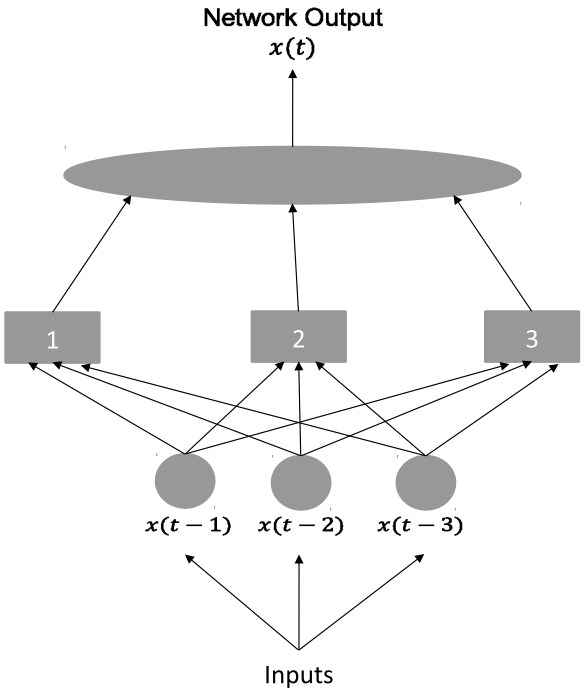


Figure 2.1: **Architecture of NARNN model:** This figure describes how NARNN works. It has three layers: input, hidden, and output. In this example, the input layer consists of three inputs units, which are the lagged values of a time series, the hidden layer consists of three hidden units, and the output layer consists of the predicted time series at period t.

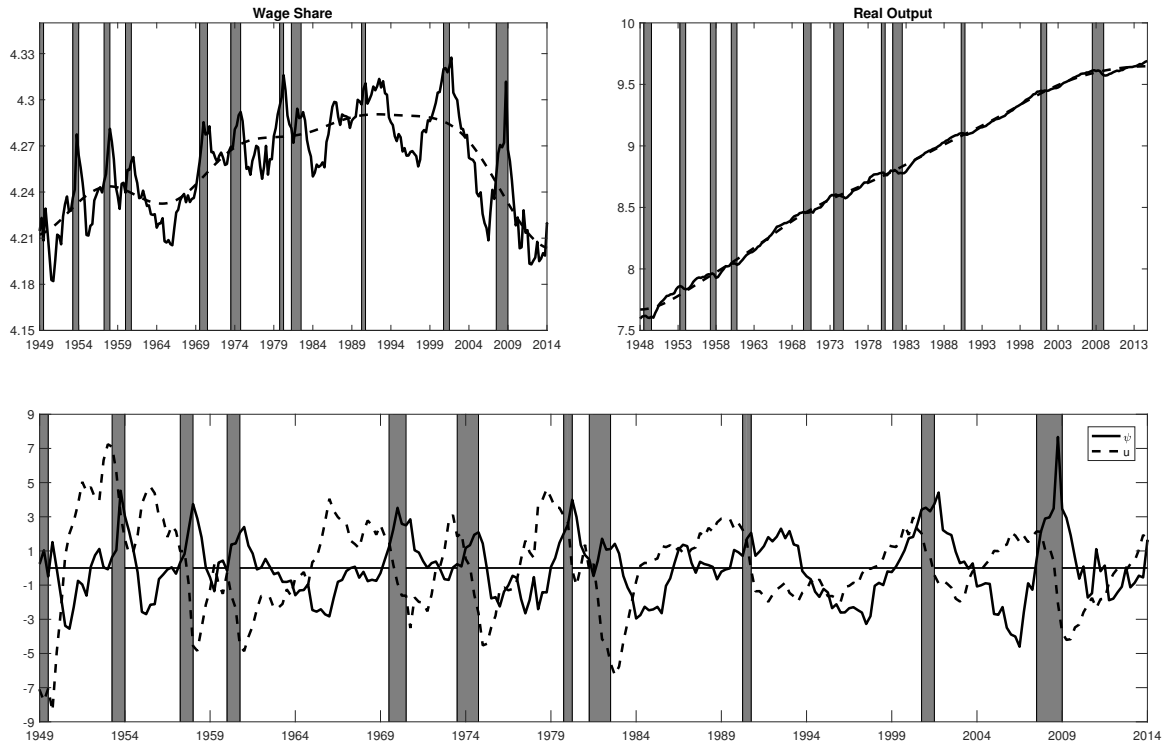


Figure 2.2: **Wage share and output gap**. Top panels: log of corporate labor cost as a share of corporate NVA and log of real output and their trends. Trend is defined as fluctuation with periodicities larger than 64 quarters per cycle. Bottom panel: detrended log wage share (ψ) and output gap (u). Both defined as percentage deviation from their trend.

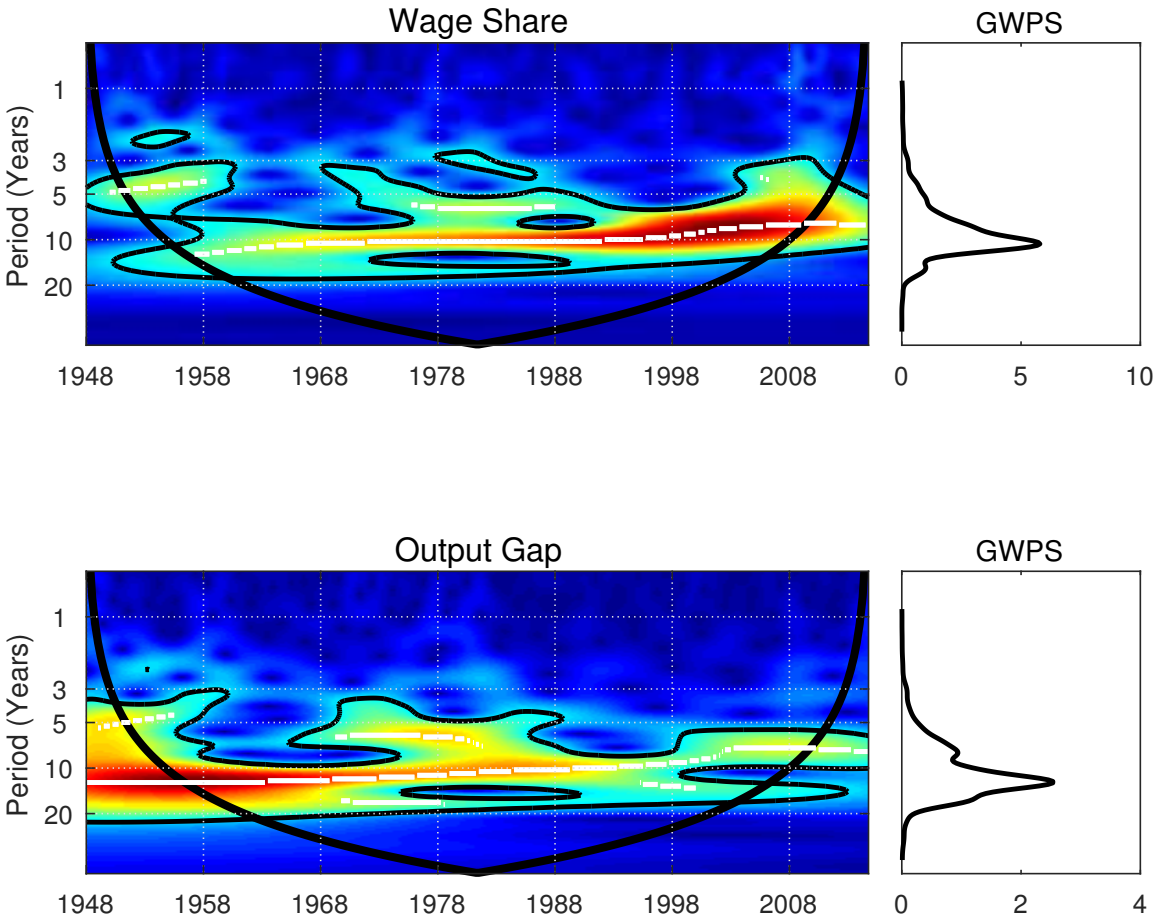
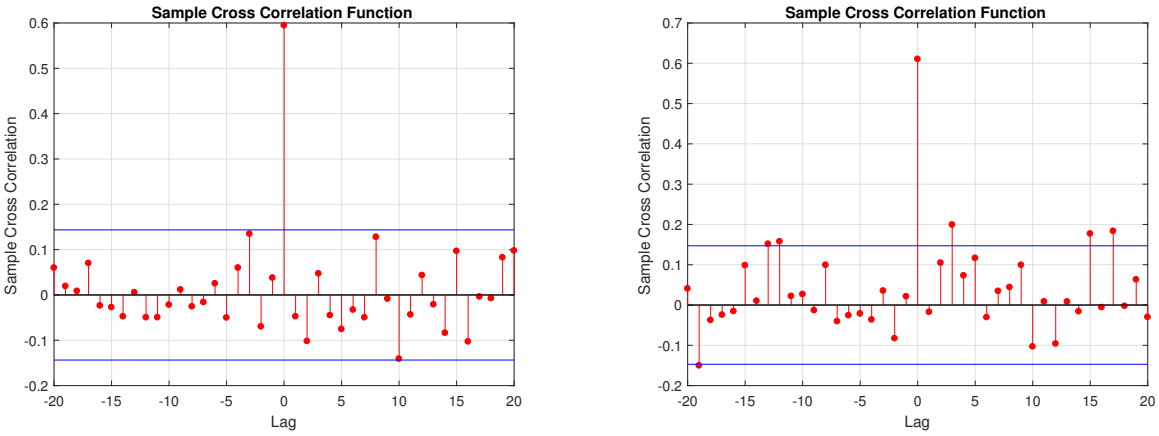


Figure 2.3: **Wavelet power spectrum and global wavelet power spectrum**: The first column displays the wavelet power spectrum for wage share and capacity utilization: warmer colors represent larger power, i.e., red regions depict the largest and blue the lowest. White lines display the largest power at each time, the so–called 'characteristic period.' The cone of influence that represents regions affected by edge effect is represented by a thick black line. The second column shows the global wavelet power spectrum (GWPS) that is the time average power spectrum, which resembles the Fourier power spectrum.

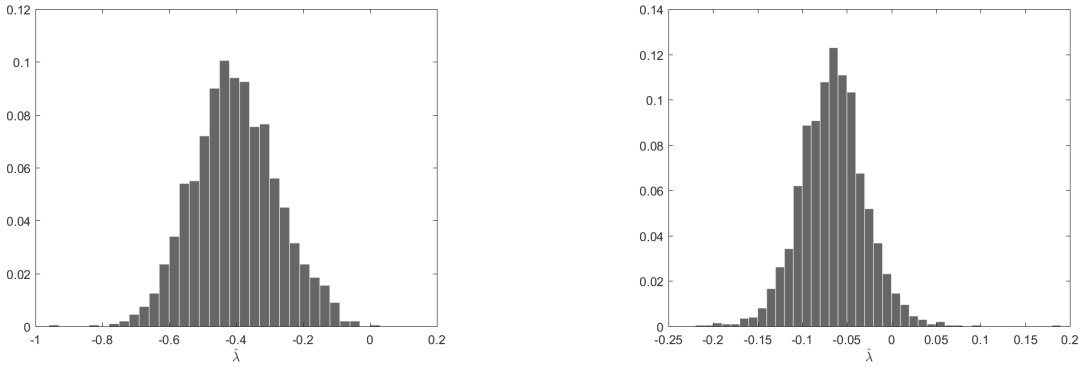
Table 2.1: **Estimation results**

	Wage Share		Output Gap	
	Hidden units (j)		Hidden units (j)	
	0	1	0	1
β_j	-0.0990	-0.8748	0.0792	1.6807
μ_j		-0.1051		-0.0368
Lag (i)		γ_{ji}		γ_{ji}
1		-1.4619		0.7187
2		0.3283		-0.0493
3		–		-0.1332
4		–		0.0488
5		–		-0.0778
6		–		0.0677
7		–		-0.0412
8		–		-0.0755
9		–		0.1277
10		–		0.0256
11		–		-0.0944
Root MSE		0.1820		0.1360
GCV(1.4)		0.0356		0.0231
R^2		88.70%		93.93%
LE		-0.3607		-0.0657
LE CI($\alpha = 0.1$)		(-0.6010,-0.1820)		(-0.1250 ,-0.0067)



(a) Wage share (b) Output gap

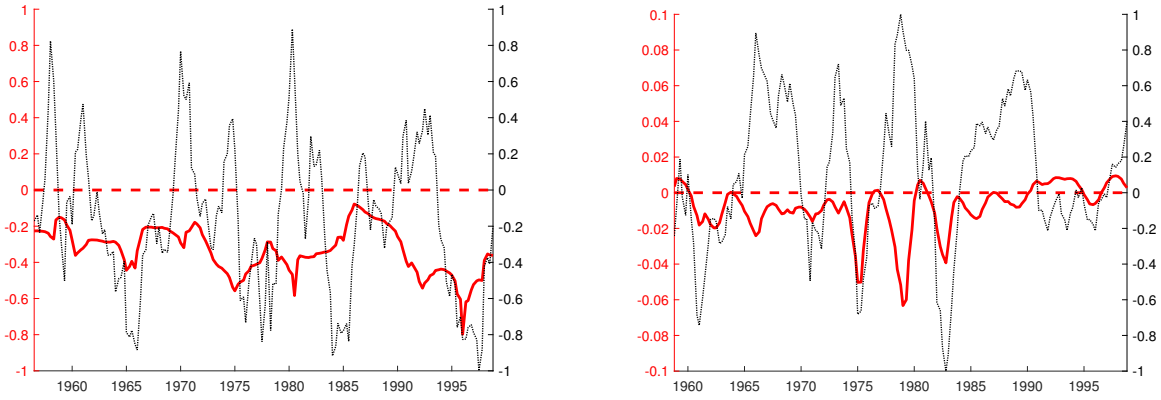
Figure 2.4: **Analysis of residuals:** These figures show the correlation analysis in equation (2.13) for wage share (left panel) and output gap (right panel).



(a) Wage share

(b) Output gap

Figure 2.5: **Dominant Lyapunov exponents bootstrap distribution:** These figures illustrate the distribution of the dominant Lyaunov exponent, which is used to calculate the confidence interval, at 10% of significance, in Table 2.1, respectively. See Appendix B.3 for further details on bootstrapping.



(a) Wage share

(b) Output gap

Figure 2.6: **Local Lyapunov exponent (LLE):** These figures present the local Lyapunov exponent (LLE) (red lines) for $\omega = 25$ quarters (see subsection 2.3.1 for further details). Left panel for wage share and right panel output gap (dotted lines).

CHAPTER 3

ENDOGENOUS FLUCTUATIONS IN DEMAND AND DISTRIBUTION: AN EMPIRICAL INVESTIGATION

3.1 Abstract

This paper empirically investigates the possibility of self–sustained oscillations at business cycle frequency between aggregate demand and the functional income distribution. Using wavelet decompositions, we identify cyclical variations of GDP and wage share from their respective long–run trends in the US, 1956–2004. To allow for identification of nonlinear dynamic interaction, we employ a black–box neural network approach called *Nonlinear AutoRegressive Neural Network* (NARNN). The method is first tested on simulated data with added random disturbances with and without nonlinearities in the output gap and distributive isoclines; NARNN correctly identifies the underlying models. Results for US data indicate the existence of a stable limit cycle with wage share leading output gap in a cycle à la Goodwin (1967). Further, NARNN suggests that the local instability that drives the limit cycle might be located in the own–feedback of the wage share. This is in contrast to other investigations that have put forth hypotheses about either a locally unstable accelerator in demand, or a nonlinearity in the cross–feedback from an activity variable to the wage share.

3.2 Introduction

Limit cycle oscillations are not new and have been the focus of countless research efforts. Seminal contributions include Kalecki (1937), Kaldor (1940), and Goodwin (1951, 1967). Generally, in these models, the capitalist system is depicted as one in which the interaction among agents generates perpetual, self–sustaining oscillations. An important implication is that booms and recessions are connected and are part of the attractor of the dynamical system. Contrary to this view stands the research agenda started by Frisch

(1933) on *equilibrium models*.[1] According to this line of thought, the macroeconomy is stable and fluctuations occur due to shocks disturbing the steady state. The disturbances afterwards propagate through the economic system until it again settles down. Therefore, booms and recession are disconnected and transient states. Modern mainstream models such as Real Business Cycle (RBC) and Dynamic Stochastic General Equilibrium (DSGE) models maintain Frisch's spirit.[2]

Hand in hand with this research agenda, some time series econometric tools have evolved in order to give it empirical meaning (see for instance Qin and Gilbert (2001)). The crucial difference is whether the error term is interpreted as residuals (*aberrations*) or shocks (*stimuli*). The former does not have an economic interpretation, and does not play a relevant role in the analysis of the system. The latter, in contrast, is essential to understand the sources of economic fluctuations. Important tools regarding these *stimuli* are Structural VAR (SVAR) and Forecast Error Variance Decomposition (FEVD).

This paper investigates the possibility of self–sustained oscillations, or limit cycles, in a post–Keynesian model of growth and distribution. This literature builds on Goodwin (1967), but sees the growth cycle as driven by the forces of effective demand. Seminal contributions include Barbosa-Filho and Taylor (2006) as well as Flaschel (2009, 2015). These investigations describe the interaction between functional income distribution (wage share) and effective demand (output gap) in a two–dimensional system of differential equations. Barbosa-Filho and Taylor (2006) use a VAR model to estimate the model, and find that the US economy follows a profit–led/profit–squeeze regime in demand and distribution, respectively. This implies a damped Goodwin cycle, i.e. a counter–clockwise movement in output gap and wage share.

In order to consider the empirical possibility of a limit cycle, nonlinear statistical models have to be used. In principle, this is not hard provided some prior knowledge about nonlinearities is available. That, of course, is usually not the case. For that reason, we use a model of the class of Recurrent Neural Networks (RNN) called Nonlinear AutoRegressive

---

[1] See P. Chen (2010, Chapter 12) for a detailed discussion in what he calls a *mysterious success*.

[2] Kydland and Prescott (1982) represents a starting point for RBC, and much cited DSGE papers (Christiano, Eichenbaum, & Evans, 2005; Smets & Wouters, 2007). Exceptions exist, for instance, Beaudry, Galizia, and Portier (2015, 2016a, 2016b).

Neural Network (NARNN). A key feature of this approach is that it does not require linearization.

Our results indicate that the interaction between wage share and output gap features a stable limit cycle à la Goodwin. Furthermore, *local* instability arises from wage share own–feedback.

The remainder of the paper is organized as follows: section 3.3 briefly describes the structuralist model of effective demand and income distribution. Section 3.4 presents the data, empirical methods and related tests. Section 3.5 reports simulation and estimation results. The last section concludes.

3.3 A model in demand and distribution

The structuralist model for effective demand and income distribution has two state variables: functional income distribution and level of economic activity. Usually, the former is the wage share (ψ), defined as $\psi = W/P\xi$ where W, P, and ξ are nominal wage rate, price level, and labor productivity, respectively. The profit share follows as $\pi = 1 - \psi$. Activity might be defined as the rate of capacity utilization (or output gap) $U = Y/Y^*$, which is often proxied by the observed output–captial ratio, $u = U\sigma$ where $Y, Y^*, \sigma = Y^*/K$ and K are real output, potential output, potential output to capital ratio, and capital stock, respectively.

On the distributive side, the growth rate of the wage share can be written as

$$\hat{\psi} = \hat{W} - \hat{P} - \hat{\xi} = f^W(u, \psi) - f^P(u, \psi) - f^{\xi}(u, \psi), \tag{3.1}$$

with $\hat{x} = \dot{x}/x$ and $f^j, j = \{W, P, \xi\}$ the corresponding behavioral functional forms. Equation (3.1) provides a fairly general functional form for the wage share dynamics. Different articles focus on different parts of equation (3.1). For instance, Flaschel and Krolzig (2006), Proaño, Flaschel, Ernst, and Semmler (2006), and Tavani, Flaschel, and Taylor (2011) study the wage–price spiral by exploring the *wage*–Phillips Curve (f^W) and *price*–Phillips Curve (f^P). These models usually assume that the own feedback of wage share, i.e., $\partial\hat{\psi}/\partial\psi$, is stable and $\partial\hat{\psi}/\partial u$ is ambiguous. If positive (negative), the distributive regime is called profit–squeeze or labor market-led real wage adjustment (forced savings or goods market-led real wage adjustment).

Further, the growth rate of labor productivity, $\hat{\xi} = f^{\xi}(u, \psi)$, might be endogenous, raising the possibility of ambiguity on the sign of $\partial\hat{\psi}/\partial\psi$. See, for instance Taylor (2004, Chapter 7). Let $f^{\omega} = f^{W} - f^{P}$, so that wage share own feedback becomes $f^{\omega}_{\psi} - f^{\xi}_{\psi}$. Instability might arise if real wages respond *positively* to the wage share, or if labor productivity responds *negatively* to the labor share.

The level of economic activity is assumed to be determined by effective demand. Here, we define demand as the income–capital ratio, proxied as $u = U\sigma$. (In the empirical application, we obtain a proxy for U by detrending real GDP.) Its growth rate, assuming constant σ, is given by

$$\hat{u} = \hat{Y} - \hat{K} = g^{Y}(u, \psi) - g^{K}(u, \psi), \tag{3.2}$$

with g^{Y} and g^{K} output growth and accumulation function (see as well Skott (1989)). As above, g^{j}_{i} stands for the partial derivative of g^{j} with respect to i. It is assumed that $g^{K}_{u} > 0$. If $g^{Y}_{u} < 0$, the goods market adjustment is stable. Skott (1989), Flaschel (2009, 2015), and von Arnim and Barrales (2015) investigate theoretically the possibility of a Goodwin–type limit cycle driven by a nonlinear activity process, which would here imply $g^{Y}_{u} - g^{K}_{u} > 0$. Further, if $g^{Y}_{\psi} - g^{K}_{\psi} < 0$, the regime is called *profit–led*, if $g^{Y}_{\psi} - g^{K}_{\psi} > 0$, *wage–led*.

The dynamical system can be summarized as

$$\begin{aligned} \hat{u} &= g^{Y}(u, \psi) - g^{K}(u, \psi) \\ \hat{\psi} &= f^{W}(u, \psi) - f^{P}(u, \psi) - f^{\xi}(u, \psi), \end{aligned} \tag{3.3}$$

with the Jacobian matrix at the steady state, J^{*}, given by

$$J^{*}(u, \psi) = \begin{pmatrix} g^{Y}_{u} - g^{K}_{u} & g^{Y}_{\psi} - g^{K}_{\psi} \\ f^{W}_{u} - f^{P}_{u} - f^{\xi}_{u} & f^{W}_{\psi} - f^{P}_{\psi} - f^{\xi}_{\psi} \end{pmatrix}. \tag{3.4}$$

The dynamical system in (3.3) can generate different dynamics depending on the assumptions imposed. Our prior builds on a profit–squeeze/profit–led regime, which the available evidence supports . For detailed discussions, see Barbosa-Filho and Taylor (2006), Mohun and Veneziani (2008), Kiefer and Rada (2015), and Barrales and von Arnim (2017). In the next section, we empirically investigate the reduced form sign of this system without linearization.

3.4 Data and estimation

This section describes the data employed in the estimation, the estimation procedure—the multivariate NARNN—and related tests on model validity.

3.4.1 Data

The state variables in the system (3.3) are the output gap and wage share. Wage share is computed as the corporate labor cost as a share of corporate net value added from BEA NIPA table 1.14. It is available quarterly from 1948:I to 2014:IV. A detailed discussion can be found in Barrales and von Arnim (2017). Capacity utilization is measured as the output gap. Figure 3.1 top panels show the log of wage share and real output and their trends. Figure 3.1 bottom panel presents output gap and detrended wage share measured as the percentage deviation from their trends.

Trend is measured as periodicities longer than 64 quarters per cycle. In this, we follow Comin and Gertler (2006) and Pancrazi (2015), who have shown that medium–run dynamics longer than 32 quarters per cycle—usually considered as a trend—greatly influence business cycles fluctuations. The trend is then calculated with the Maximal Overlap Discrete Wavelet Transform (MODWT) (see for instance, Percival and Walden (2000) and Appendix B.1). In a nutshell, MODWT for level J for a time series $\boldsymbol{X}$ yields highly redundant and non–orthogonal column vectors $\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \cdots, \tilde{\boldsymbol{W}}_J$ and $\tilde{\boldsymbol{V}}_J$ each of dimension T. Vectors $\tilde{\boldsymbol{W}}_j$ are associated with an approximated bandpass filter with frequencies between $[2^{-(j+1)}, 2^{-j}]$ (or periodicities between $[2^j, 2^{j+1}]$ quarters per cycle); meanwhile $\tilde{\boldsymbol{V}}_J$ is associated with a low–pass filter with approximate frequencies $[0, 2^{-(J+1)}]$ (or periodicities larger than 2^{J+1} quarters per cycles).

MODWT yields an energy decomposition

$$||\boldsymbol{X}||^2 = \sum_{j=1}^{J} ||\tilde{\boldsymbol{W}}_j||^2 + ||\tilde{\boldsymbol{V}}_J||^2, \tag{3.5}$$

where $||\bullet||$ is the l^2–norm, and an additive decomposition called Multi–Resolution–Analysis (MRA)

$$\boldsymbol{X} = \sum_{j=1}^{J} \tilde{\boldsymbol{\mathcal{D}}}_j + \tilde{\boldsymbol{\mathcal{S}}}_J, \tag{3.6}$$

where $\tilde{\boldsymbol{\mathcal{D}}}_j$ and $\tilde{\boldsymbol{\mathcal{S}}}_J$ are the j–th order *detail* and the J–th order *smooth* for $\boldsymbol{X}$, respectively. The transform from $\boldsymbol{X}$ to $\tilde{\boldsymbol{W}}_j$ and from $\boldsymbol{X}$ to $\tilde{\boldsymbol{V}}_j$ can be expressed as

$$\tilde{\boldsymbol{W}}_j = \tilde{\mathcal{W}}_j \boldsymbol{X} \quad \text{and} \quad \tilde{\boldsymbol{V}}_j = \tilde{\mathcal{V}}_j \boldsymbol{X}, \tag{3.7}$$

where each row of the $N \times N$ matrix $\tilde{W}_j$ has a value of the periodized filter to length T of $\{\tilde{h}_j\}$ and $\{\tilde{g}_j\}$, which are called wavelet and scaling filters, respectively.

The MRA is obtained, therefore, as

$$\tilde{\boldsymbol{\mathcal{D}}}_j \equiv \tilde{\mathcal{W}}_j^T \tilde{\boldsymbol{W}}_j \quad \text{and} \quad \tilde{\mathbf{S}}_j \equiv \tilde{\mathcal{V}}_j^T \tilde{\boldsymbol{V}}_j, \tag{3.8}$$

where $\tilde{\mathbf{S}}_J$ is defined as the trend of $\boldsymbol{X}$. We choose the Daubechies' least asymmetric (LA) wavelet filter LA(8) and $J = 5$.

Before the training process (estimation), we assess the possibility of limit cycles in both variables through the examination of their spectral properties. Figure 3.2 presents the Continuous Wavelet Transform (CWT) power spectrum, which is a timescale localized spectral density, and the Global Wavelet Power Spectrum (GWPS), which averages the spectrum along time (see Appendix B.2 for further details, specifically the definition in equation (B.18)). Taking in consideration that both variables are detrended, i.e., higher frequencies are not removed, output gap and wage share are clearly band limited; therefore, a limit cycle behavior seems likely.

3.4.2 Estimation

System identification can be defined as the process in which a mathematical model is used in order to map experimental data by minimizing some performance index between data and system output. Since the seminal work of Narendra and Parthasarathy (1990, 1992), Artificial Neural Network has been used successfully for nonlinear system identification. Noël and Kerschen (2017) provide a recent review of the state of the art. The estimation is carried out equation by equation using Nonlinear AutoRegressive Neural Network (NARNN) (see S. Chen, Billings, & Grant, 1990; Narendra & Parthasarathy, 1990), which can be written as follows:

$$\underbrace{\hat{\boldsymbol{y}}(t)}_{n\times 1} = \underbrace{\beta_0}_{n\times 1} + \sum_{j=1}^{K} \underbrace{\beta_{1j}}_{n\times 1} g\Big(\underbrace{\gamma_{0j}}_{n\times 1} + \underbrace{\gamma_j}_{n\times np_y} \underbrace{\boldsymbol{Y}^*(t-1)}_{np_y\times 1} \Big), \tag{3.9}$$

where n stands for the number of endogenous variables, p_y number of lags, $g(\bullet)$ the activation function which is chosen to be $\tanh(\bullet)$, K number of hidden layers, β_0 and γ_{0j}

biases, β_{1j} and γ_j weights, $\hat{\boldsymbol{y}}(t)$ output neurons, $\boldsymbol{Y}^*(t-1)$ inputs neurons, and $\boldsymbol{y}(t)$ targets. Targets, outputs, and input neurons are arranged as

$$\begin{aligned}
\boldsymbol{y}(t) &= [y_1(t), y_2(t), \cdots, y_n(t)]^T, \\
\hat{\boldsymbol{y}}(t) &= [\hat{y}_1(t), \hat{y}_2(t), \cdots, \hat{y}_n(t)]^T, \\
\boldsymbol{Y}^*(t-1) &= \left[y_1^*(t-1), \cdots, y_n^*(t-1); y_1^*(t-2), \cdots, y_n^*(t-2); \cdots; y_1^*(t-p_y), \cdots, y_n^*(t-p_y)\right]^T
\end{aligned}$$

with T as matrix transpose.

There are two model configurations: series–parallel (or open loop), $\boldsymbol{Y}^*(t) = \boldsymbol{Y}(t)$, and parallel (or closed loop), $\boldsymbol{Y}^*(t) = \hat{\boldsymbol{Y}}(t)$, models. We employ series–parallel for training and parallel for simulation. The former has the advantage that a common feedforward neural network can be used. Furthermore, K and p_y define the complexity of the model.

For training and in order to reduce the possibility of overfitting, Bayesian regularization plus cross–validation are employed. In a nutshell, Bayesian regularization constrains the *growth* of the size of the parameters instead of the *number* of parameters as with a typical information criteria. The performance index to be minimized becomes:

$$F(\boldsymbol{\omega}) = \beta E_D + \alpha E_W, \tag{3.10}$$

where $\boldsymbol{\omega} \in \mathbb{R}^m$ vector of network weights (weights and biases), E_W sum of square network weights, E_D mean square error ($MSE = T^{-1}\sum_{t=1}^{T}[\hat{y}(t) - y(t)]^2$), and α/β controls the effective complexity of the network solution, i.e., the larger the ratio, the smoother the network response (see Appendix B.6 for a detailed explanation).

A crucial step for system identification is cross–validation. The key issue is that the residuals should be white noise after the appropriate model is selected. For linear models, the goal is that residuals should be uncorrelated with past inputs, outputs, and residuals. This is accomplished by estimating the residual autocorrelation and cross–correlation with inputs. However, for nonlinear models, this is not enough: passing such tests, the residuals may still contain nonlinear terms. The main concept in nonlinear model validation is that residuals should be unpredictable from *all linear and nonlinear combinations* of past inputs, outputs, and residuals; see Billings and Voon (1983). We therefore use the test introduced by Billings and Zhu (1995) in a Multiple Input Multiple Output (MIMO) context. However, since there are no exogenous inputs in this application, testing is reduced to:

$$\phi_{\eta\gamma}(\tau) = \lambda_0 \delta(\tau), \quad \forall\tau \tag{3.11}$$

where δ is the delta function, $\lambda_0 = \sqrt{\sum_{t=1}^{T} (\eta(t) - \bar{\eta})^2} \Big/ \sqrt{\sum_{t=1}^{T} (\gamma(t) - \bar{\gamma})^2}$ is a constant, and η and γ are defined as

$$\begin{aligned} \eta(t) &= \epsilon_1(t)^2 + \epsilon_2(t)^2 + \cdots + \epsilon_n(t)^2, \\ \gamma(t) &= y_1(t)\epsilon_1(t) + y_2(t)\epsilon_2(t) + \cdots + y_n(t)\epsilon_n(t), \end{aligned} \tag{3.12}$$

with $\epsilon_i(t)$ the residual term for period t and equation i. The 95% confidence bands are approximately $\pm 1.96/\sqrt{T}$ for T length of the data. $\phi_{xy}(\bullet)$ is the cross–correlation function between x and y, i.e.,

$$\phi_{xy}(\tau) = \frac{\sum_{t=1}^{N-\tau} [x(t) - \bar{x}] [y(t+\tau) - \bar{y}]}{\sqrt{\sum_{t=1}^{N} [x(t) - \bar{x}]^2} \sqrt{\sum_{t=1}^{N} [y(t) - \bar{y}]^2}}. \tag{3.13}$$

3.5 Empirical results

This section first illustrates the multivariate NARNN on the basis of two data sets, generated on different assumptions about the underlying Goodwin–Kalecki model. We then report results for the US detrended data previously discussed, and close with an analysis of the resulting isoclines.

3.5.1 Simulations

Before the next subsection reports estimation results, we provide two controlled experiments to assess the ability of NARNN to identify both *dynamics* and *isoclines* in simulated models of effective demand and functional income distribution. We build here on work presented in von Arnim and Barrales (2015).

Following that approach, the dynamics of the two–dimensional system in capacity utilization and wage share can be written as

$$\begin{aligned} \dot{u} &= \alpha \left(h_u \tan^{-1}(u - \bar{u}) - g_u(u - \bar{u}) + \mu_\psi(\psi - \bar{\psi}) \right) \\ \dot{\psi} &= \beta \left(a_u(u - \bar{u}) + a_\psi(\psi - \bar{\psi}) \right) \end{aligned} \tag{3.14}$$

where $h_u, g_u, a_u > 0$ and $a_\psi, \mu_\psi < 0$. A supercritical Hopf bifurcation exist when h_u passes through $h_u^0 = g_u - (\alpha/\beta)a_\psi$.

Figures 3.3 and 3.4 show the *Stable* and *Limit Cycle* Goodwin–Kalecki models. Both systems are simulated by using the Euler-Muruyama approximation scheme with time

step $\delta t = 0.01$ and standard Brownian motion (see for instance, Malham and Wiese (2010)). Then 300 observations are sampled with sampling rate $1/\delta t$. Table 3.1 shows parameters employed. The sign patterns of the Jacobian matrices are

$$J^{\text{stable}}(u,\psi) = \begin{pmatrix} - & - \\ + & - \end{pmatrix} \quad \text{and} \quad J^{\text{limit cycle}}(u,\psi) = \begin{pmatrix} \pm & - \\ + & - \end{pmatrix}. \tag{3.15}$$

NARNN is trained for both with normalized data between -1 and 1. Results on the dynamics and isoclines are presented in Figures 3.5 and 3.6, with $p_y = 1$ and $K = 2$. NARNN correctly identifies the *skeleton* of the models, as well as their isoclines and associated signs of the Jacobian matrix in equations (3.15).

3.5.2 NARNN results

The sample selected goes from 1956:I to 2004:IV and each time series is normalized between -1 and 1 by

$$x_n = \frac{2(x - x_{\min})}{(x_{\max} - x_{\min})} - 1 \tag{3.16}$$

where x, x_n, $x_{\min}$, $x_{\max}$ are the original, normalized, maximum, and minimum data, respectively.

The first step is to choose the smallest model that passes the test in equation (3.11). Figure (3.7) bottom panel shows $K = 2$ and $p_y = 1$ is enough to eliminate any sign of either linear or nonlinear structure in the residuals.

With $n = 2$, $\boldsymbol{y}(t) = (\psi(t), u(t))^T$, ψ wage share, u output gap, $\boldsymbol{Y}(t) = (\psi(t-1), u(t-1))^T$, and network weights and biases (see equation (3.9)):

$$\beta_0 = (-0.0337, 0.0764)^T, \tag{3.17}$$

$$\beta_{11} = (-0.9190, -0.6101)^T \quad \& \quad \beta_{12} = (0.9706, -0.4577)^T, \tag{3.18}$$

$$\gamma_{01} = (-0.0820, 0.1022)^T \quad \& \quad \gamma_{02} = (-0.0860, 0.0208)^T, \tag{3.19}$$

$$\gamma_1 = \begin{pmatrix} -1.1172 & 0.0893 \\ 0.2880 & 0.8762 \end{pmatrix} \quad \& \quad \gamma_2 = \begin{pmatrix} -0.1045 & -0.4656 \\ 1.0012 & -0.2695 \end{pmatrix}, \tag{3.20}$$

where T stands for matrix transpose.

After training and cross–validation, the resulting NARNN model is simulated using the parallel configuration. Figure 3.7 top panel presents the simulation results, which clearly exhibit a limit cycle. Furthermore, wage share leads capacity utilization in a counter–clockwise movement. This is the empirically relevant pattern; see Barbosa-Filho and Taylor (2006), Mohun and Veneziani (2008), Kiefer and Rada (2015), and Barrales and von

Arnim (2017). Figure 3.7 middle panel compares the GWPS of the original data and the simulated model. The spectral densities of original and NARNN–generated data show similarity.

3.5.3 What produces the limit cycles?

Our results in the last subsection suggest that output gap and wage share at business cycle frequency in the US can generate a self–sustaining oscillation. What produces this disequilibrium dynamics?

Provided that the network is small enough, this question can be addressed using the estimated system (3.9). The differential equations implied in the neural network can be written as

$$\begin{array}{rcl} \dot{u}(t) & = & -u(t) + \beta_0^u + \sum_{j=1}^{K} \beta_{1j}^u \tanh\left(\gamma_{0j}^u + \gamma_{1j}^{\psi}\psi(t) + \gamma_{1j}^u u(t)\right) \\ \dot{\psi}(t) & = & -\psi(t) + \beta_0^{\psi} + \sum_{j=1}^{K} \beta_{1j}^{\psi} \tanh\left(\gamma_{0j}^{\psi} + \gamma_{2j}^{\psi}\psi(t) + \gamma_{2j}^u u(t)\right), \end{array} \tag{3.21}$$

with Jacobian matrix

$$J(u,\psi) = \begin{pmatrix} \sum_{j=1}^{K} \beta_{1j}^u \gamma_{1j}^u \left(\frac{\partial \tanh(a_j^u)}{\partial a_j^u}\right) - 1 & \sum_{j=1}^{K} \beta_{1j}^u \gamma_{1j}^{\psi} \left(\frac{\partial \tanh(a_j^u)}{\partial a_j^u}\right) \\ \sum_{j=1}^{K} \beta_{1j}^{\psi} \gamma_{2j}^u \left(\frac{\partial \tanh(a_j^{\psi})}{\partial a_j^{\psi}}\right) & \sum_{j=1}^{K} \beta_{1j}^{\psi} \gamma_{2j}^{\psi} \left(\frac{\partial \tanh(a_j^{\psi})}{\partial a_j^{\psi}}\right) - 1 \end{pmatrix} \tag{3.22}$$

where $\frac{\partial \tanh\left(a^j\right)}{\partial a^j} \in [0,1]$ and $j = \{u, \psi\}$. Using the parameters in equations (3.17)–(3.20), the Jacobian has the following pattern

$$J^*(u,\psi) = \begin{pmatrix} - & - \\ + & \pm \end{pmatrix}. \tag{3.23}$$

Our results suggest that limit cycles are produced because of *local instability* on the wage share own–feedback. As discussed in section 3.3, this is possible if real wage growth responds strongly and positively to the wage share, or labor productivity growth reacts strongly and negatively to the wage share.

3.6 Concluding remarks

This paper provides new insights on the complex dynamics between effective demand and functional income distribution. The dynamics are explored using the multivariate Nonlinear AutoRegressive Neural Network (NARNN) model, which does not requiring linearization. The small size of the estimated neural network allows the analytical explo-

ration of its isoclines, which in turn suggests that the local instability is related to wage share's own–feedback.

Further research is needed to improve our understanding of the dynamics of the components of the wage share.

3.7 References

Barbosa-Filho, N., & Taylor, L. (2006). Distributive and demand cycles in the US economy – a structuralist Goodwin model. *Metroeconomica*, *57*(3), 389–411.

Barrales, J., & von Arnim, R. (2017). Longer run distributive cycles: Wavelet decompositions for US, 1948-2011. *Review of Keynesian Economics*, *5*(2), 196–217.

Beaudry, P., Galizia, D., & Portier, F. (2015). *Reviving the limit cycle view of macroeconomic fluctuations.* NBER Working Paper No. 21241.

Beaudry, P., Galizia, D., & Portier, F. (2016a). *Is the macroeconomy locally unstable and why should we care?* NBER Working Paper No. 22275.

Beaudry, P., Galizia, D., & Portier, F. (2016b). *Putting the cycle back into business cycle analysis.* NBER Working Paper No. 22825.

Billings, S., & Voon, W. (1983). Structure detection and model validity tests in the identification of nonlinear systems. *IEE Proceedings D - Control Theory and Applications*, *130*(4), 193–199.

Billings, S., & Zhu, Q. (1995). Model validation tests for multivariable nonlinear models including neural networks. *International Journal of Control*, *62*(4), 749-766.

Chen, P. (2010). *Economic complexity and equilibrium illusion: Essays on market instability and macro vitality* (1st ed.). New York and London: Routledge.

Chen, S., Billings, S., & Grant, P. (1990). Non–linear system identification using neural networks. *International Journal of Control*, *51*(6), 1191-1214.

Christiano, L., Eichenbaum, M., & Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, *113*(1), 1–45.

Comin, D., & Gertler, M. (2006). Medium–term business cycles. *American Economic Review*, *96*(3), 523–551.

Flaschel, P. (2009). *The macrodynamics of capitalism: Elements for a synthesis of Marx, Keynes and Schumpeter*. Berlin ; London: Springer.

Flaschel, P. (2015). Goodwin's MKS system: A baseline macro model. *Cambridge Journal of Economics*, *39*(6), 1591–1605.

Flaschel, P., & Krolzig, H.-M. (2006). Wage–price Phillips curves and macroeconomic stability: Basic structural form, estimation and analysis. In C. Chiarella, R. Franke, P. Flaschel, & W. Semmler (Eds.), *Quantitative and empirical analysis of nonlinear dynamic macromodels* (Vol. 277, pp. 7–47). Wagon Lane, UK: Emerald.

Foresee, F. D., & Hagan, M. T. (1997, Jun). Gauss–newton approximation to bayesian learning. In *Proceedings of the IEEE international conference on neural networks.* doi: 10.1109/ICNN.1997.614194

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics. In *Economic essays in honour of Gustav Cassel* (pp. 171–206). London: Allen & Unwin Ltd.

Goodwin, R. (1951). The nonlinear accelerator and the persistence of business cycles. *Econometrica*, *19*(1), 1–17.

Goodwin, R. (1967). A growth cycle. In C. Feinstein (Ed.), *Socialism, capitalism and economic growth* (pp. 54–58). Cambridge, UK: Cambridge University Press.

Kaldor, N. (1940). A model of the trade cycle. *The Economic Journal*, *50*(197), 78–92.

Kalecki, M. (1937). A theory of the business cycle. *The Review of Economic Studies*, *4*(2), 77–97.

Kiefer, D., & Rada, C. (2015). Profit maximising goes global: The race to the bottom. *Cambridge Journal of Economics*, *39*(5), 1333–1350.

Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, *50*(6), 1345–1370.

MacKay, D. (1992). Bayesian interpolation. In C. R. Smith, G. J. Erickson, & P. O. Neudorfer (Eds.), *Maximum entropy and bayesian methods: Seattle, 1991* (pp. 39–66). Dordrecht: Springer Netherlands.

Malham, S. J. A., & Wiese, A. (2010). *An introduction to SDE simulation*. arXiv:1004.0646.

Mohun, S., & Veneziani, R. (2008). Goodwin cycles and the US economy, 1948-2004. In P. Flaschel & M. Landesmann (Eds.), *Mathematical economics and the dynamics of capitalism: Goodwin's legacy continued* (pp. 107–130). London: Routledge.

Narendra, K., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, *1*(1), 4-27.

Narendra, K., & Parthasarathy, K. (1992). Neural networks and dynamical systems. *International Journal of Approximate Reasoning*, *6*(2), 109-131.

Noël, J., & Kerschen, G. (2017). Nonlinear system identification in structural dynamics: 10 more years of progress. *Mechanical Systems and Signal Processing*, *83*, 2-35.

Pancrazi, R. (2015). The heterogeneous great moderation. *European Economic Review*, *74*, 207–228.

Percival, D., & Mofjeld, H. O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, *92*(439), 868-880.

Percival, D., & Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge ; New York: Cambridge University Press.

Proaño, C., Flaschel, P., Ernst, E., & Semmler, W. (2006). Gradual wage–price adjustments and keynesian macrodynamics: Evidence from the US and the euro area. *Available online under the title: Disequilibrium Macroeconomic Dynamics, Income Distribution and Wage-Price Phillips Curves: Evidence from the US and the euro area, Macroeconomic Policy Institute (IMK), IMK Working Paper*, *4*, 2006.

Qin, D., & Gilbert, C. (2001). The error term in the history of time series econometrics. *Econometric Theory*, *17*(2), 424-450.

Skott, P. (1989). Demand, effective and struggle, class and growth, cyclical. *International Economic Review*, *30*(1), 231–247.

Smets, F., & Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, *97*(3), 586–606.

Tavani, D., Flaschel, P., & Taylor, L. (2011). Estimated non-linearities and multiple equilibria in a model of distributive-demand cycles. *International Review of Applied Economics*, *25*(5), 519–538.

Taylor, L. (2004). *Reconstructing macroeconomics*. Cambridge, MA: Harvard University Press.

von Arnim, R., & Barrales, J. (2015). Demand-driven Goodwin cycles with kaldorian and kaleckian features. *Review of Keynesian Economics*, *3*(3), 351–373.

von Bremen, H. F., Udwadia, F. E., & Proskurowski, W. (1997). An efficient QR based method for the computation of lyapunov exponents. *Physica D*, *101*(1), 1–16.

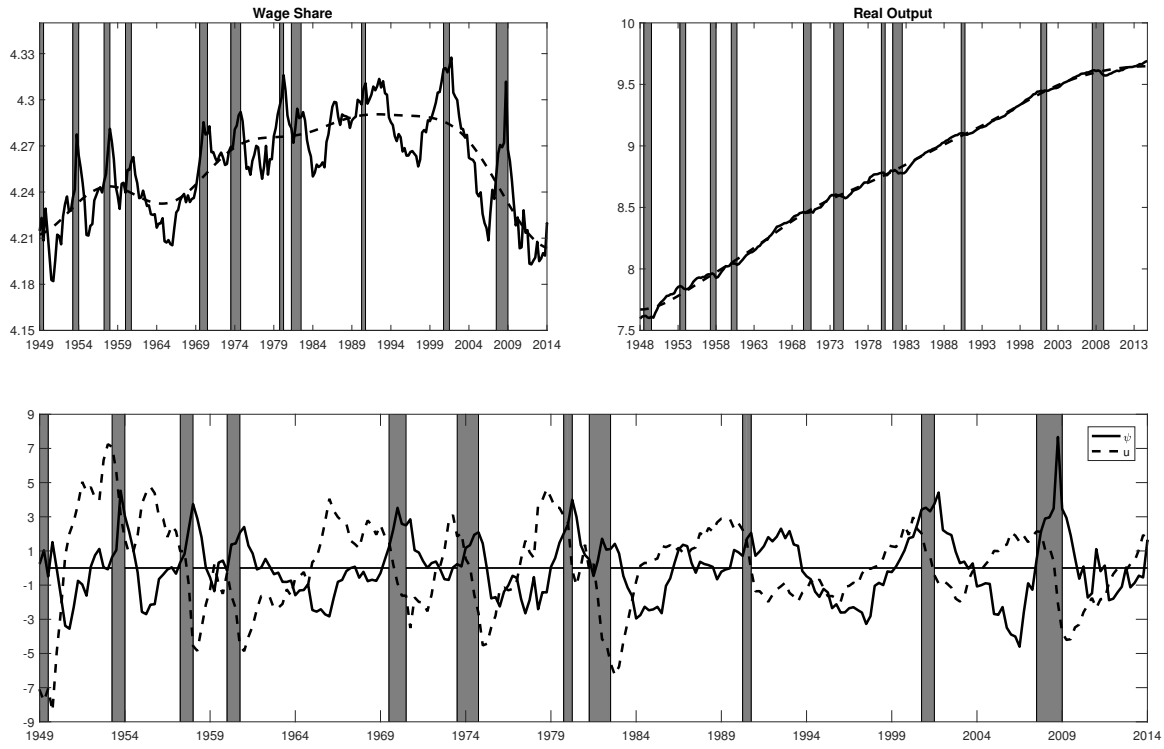


Figure 3.1: **Wage share and output gap**. Top panels: log of corporate labor cost as a share of corporate NVA and log of real output and their trends. Trend is defined as fluctuation with periodicities larger than 64 quarters per cycle. Bottom panel: detrended log wage share (ψ) and output gap (u). Both defined as percentage deviation from their trend.

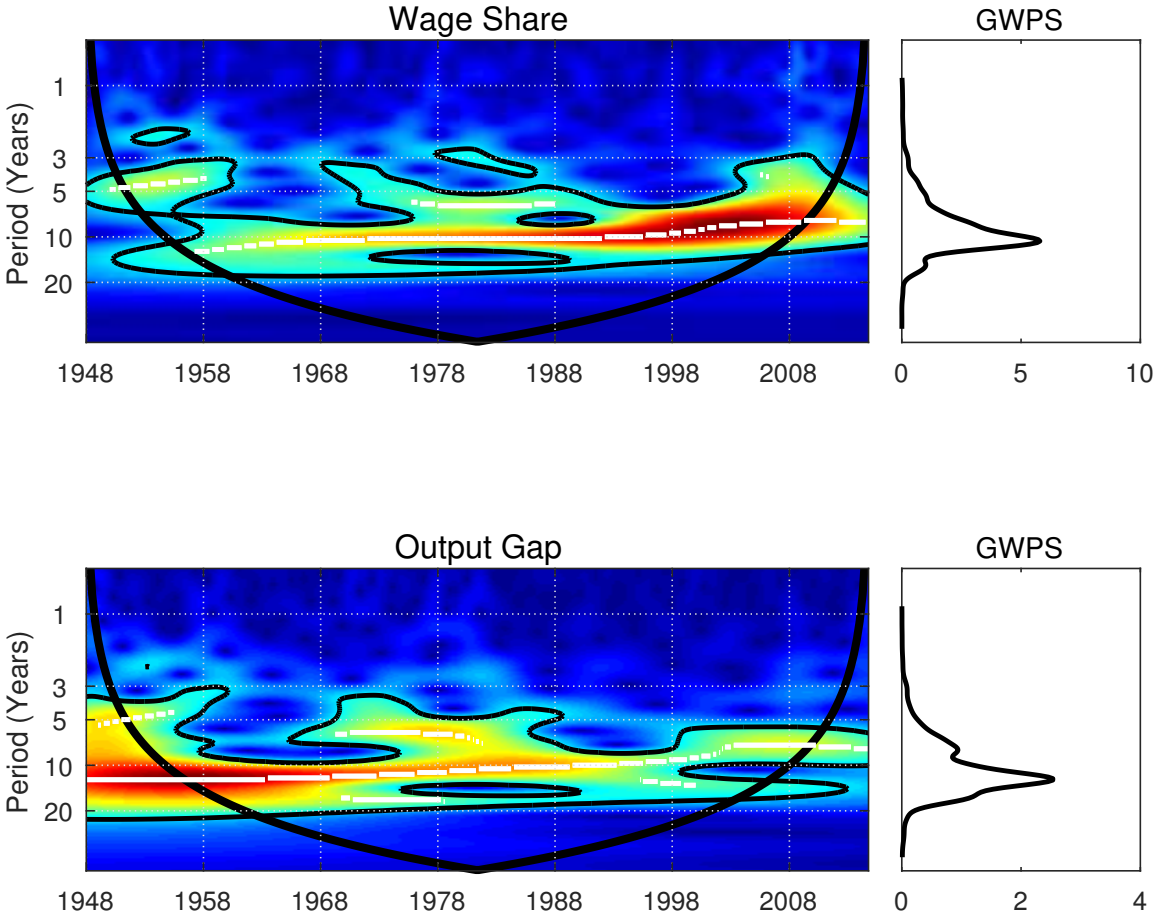


Figure 3.2: **Wavelet power spectrum and GWPS**: First column, warmer colors represent larger power, i.e., red regions depict the largest and blue the lowest. White lines display the largest power at each period, the so–called 'characteristic period'. Second column shows the Global Wavelet Power Spectrum (GWPS) which averages the spectrum along time as the Fourier transform (see equation (B.18))

Table 3.1: **Parameters:** This table presents the parameters used for simulation of the Stable Goodwin–Kalecki (Figure 3.3) and Limit Cycle Goodwin–Kalecki (Figure 3.4) models.

Parameter	α	β	g_u	μ_ψ	a_u	a_ψ	$\bar{\psi}$	$\bar{u}$	h_u^0	h_u^{stable}	$h_u^{\text{limit cycle}}$
Value	1.2	0.27	0.55	-0.3	1.4	-0.2	100	100	0.4825	0.3	1.15

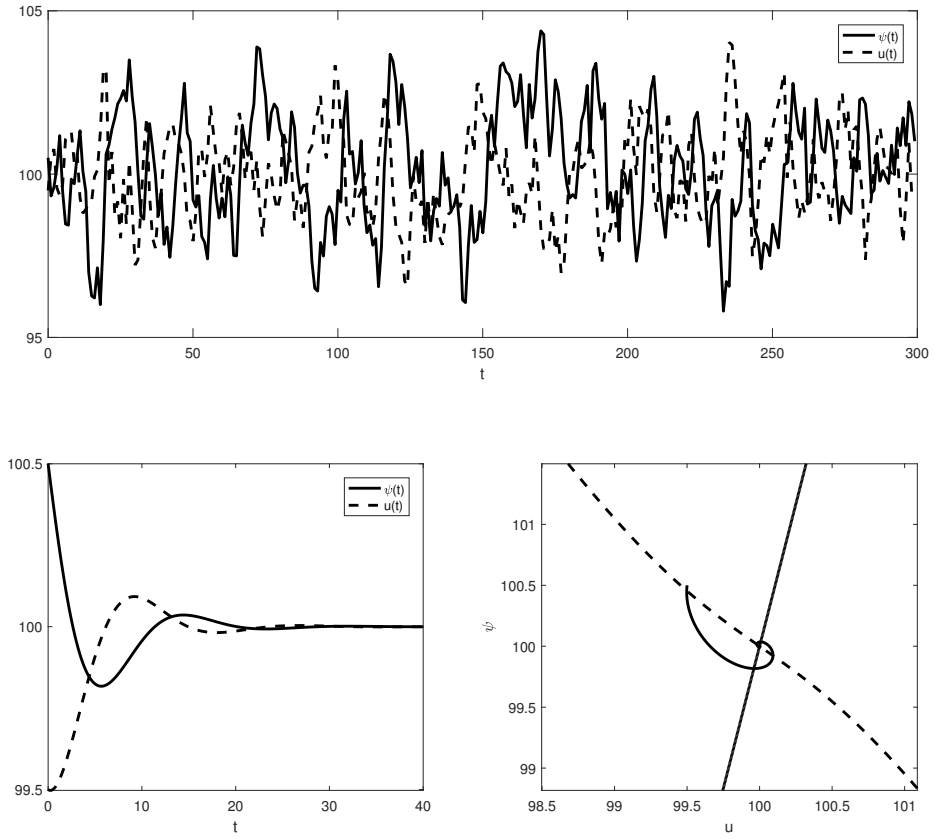


Figure 3.3: **Stable Goodwin–Kalecki:** These graphs use parameters in Table 3.1 and $h_u = 0.3 < h_u^0$ in the system (3.14). The simulation of the ordinary and stochastic differential system of equations in (3.14) is performed by the Euler-Maruyama scheme with time step $\delta t = 0.01$. Top panel presents the data employed in NARNN estimation with sampling rate $1/\delta t$. Bottom panel on the right shows the simulated "skeleton" of the model and, on the left, its phase diagram with their respective isoclines.

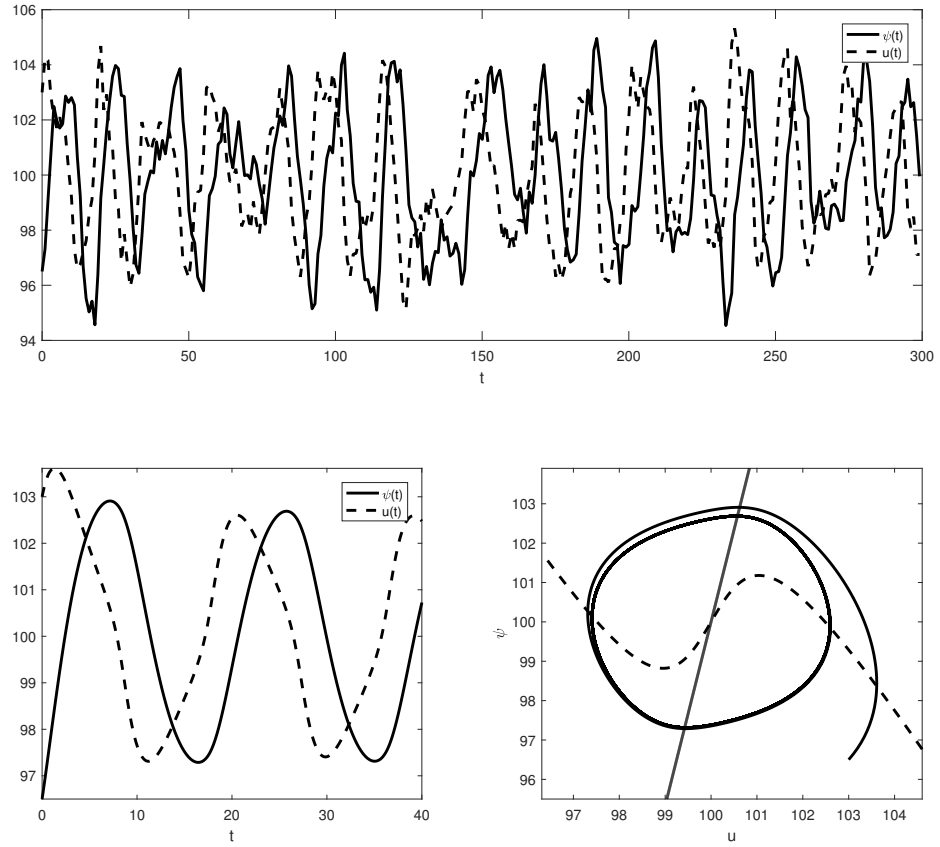


Figure 3.4: **Limit cycle Goodwin–Kalecki:** This graph uses parameters in Table 3.1 plus $h_u = 1.15 > h_u^0$ in the system (3.14). The simulation of the ordinary and stochastic differential system of equations in (3.14) is performed by the Euler-Maruyama scheme with time step $\delta t = 0.01$. Top panel presents the data employed in NARNN estimation with sampling rate $1/\delta t$. Bottom panel on the right shows the simulated "skeleton" of the model and, on the left, its phase diagram with their respective isoclines.

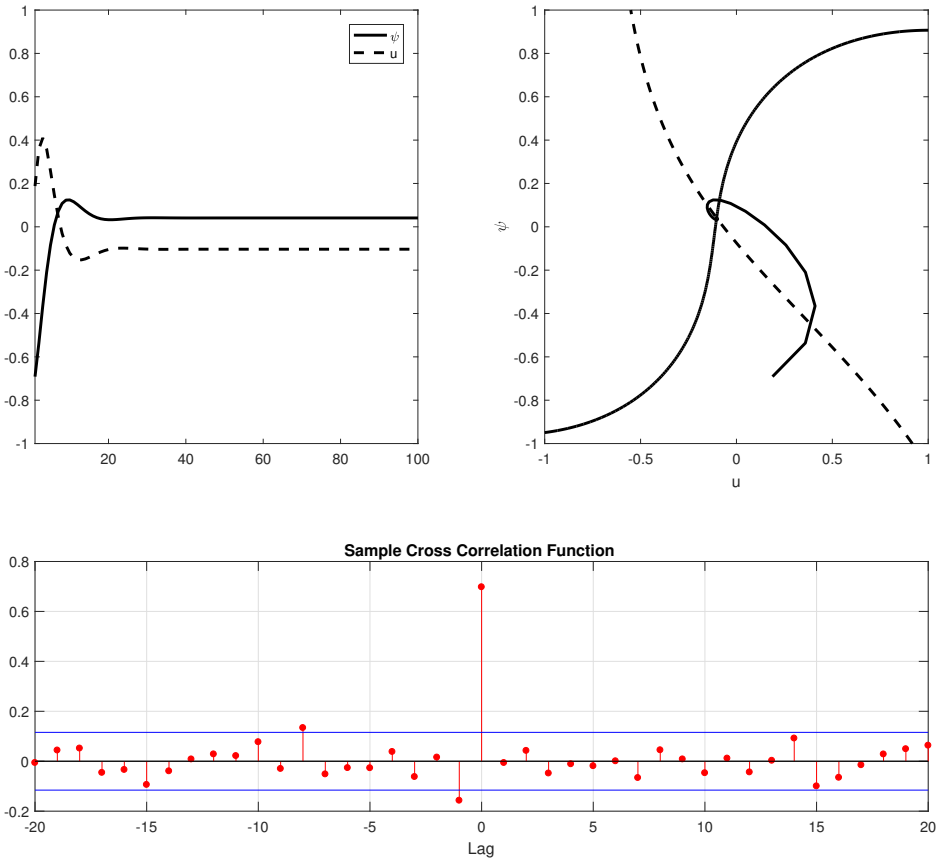


Figure 3.5: **NARNN estimation results for stable Goodwin–Kalecki:** This figure shows the results from NARNN estimation with $p_y = 1$ and $K = 2$. Top panel displays the parallel simulation, right, and their respective isoclines, left. Bottom panel shows the residual test in equation (3.11).

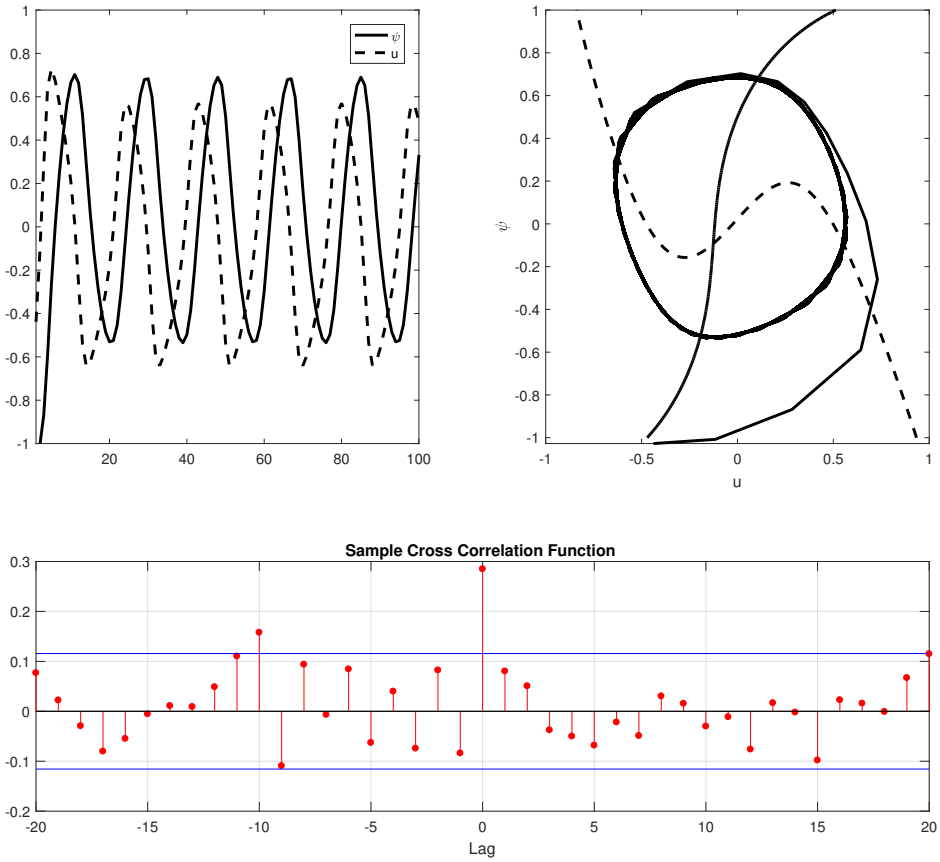


Figure 3.6: **NARNN estimation results for limit cycle Goodwin–Kalecki:** This figure shows the results from NARNN estimation with $p_y = 1$ and $K = 2$. Top panel displays the parallel simulation, right, and their respective isoclines, left. Bottom panel shows the residual test in equation (3.11).

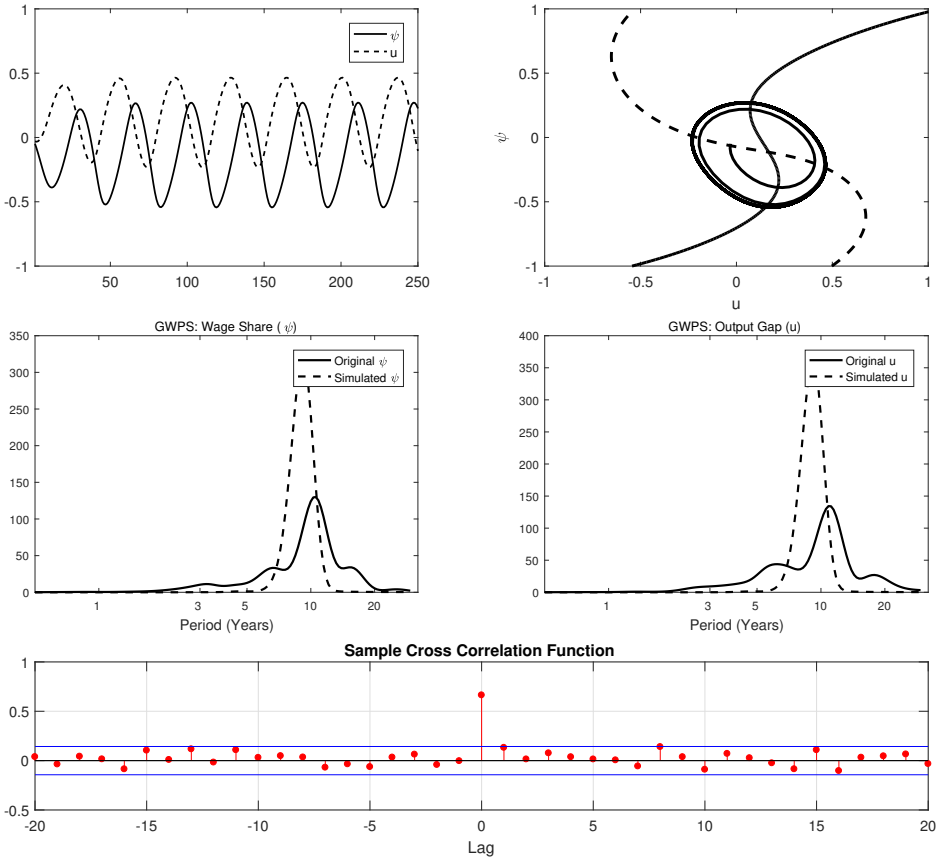


Figure 3.7: **Simulations and residual analysis:** Top left panel: simulated wage share (ψ) and output gap (u). Top right panel: the two–dimensional attractor in (u, ψ) and its isoclines continuous line for wage share and dotted for output gap. Middle panel from left to right: Global Wavelet Power Spectrum (GWPS) for the original data and simulated model, for wage share and output gap. Bottom panel, cross–validation test by cross–correlation in equation (3.11). Based on this test, the number of hidden layers and lags are found to be four and nine, respectively.

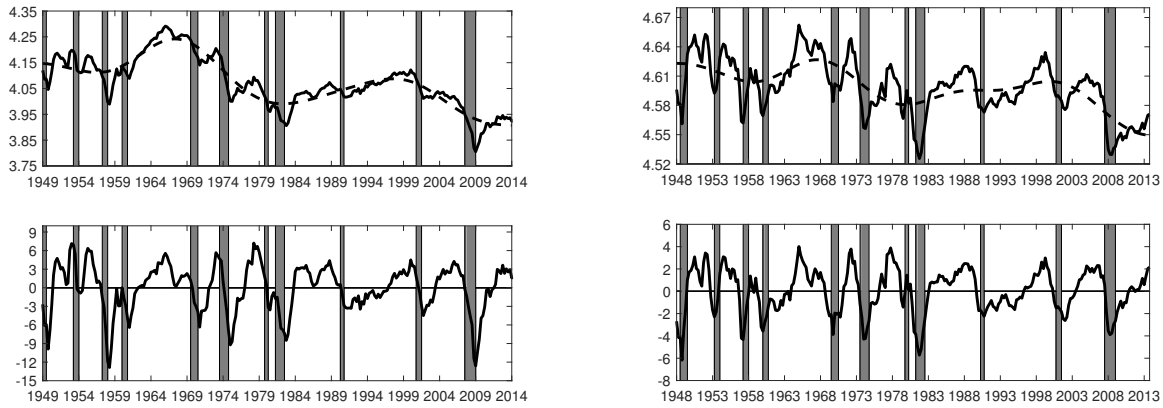
APPENDIX A

ADDITIONAL ANALYSIS: INCOME–CAPITAL RATIO AND POTENTIAL GDP

This appendix presents the estimation results for two additional measures for economic activity used in Chapter 1. They are the income–capital ratio and the real GDP to CBO's potential GDP. Income–capital ratio, see Figure A.1 first column, is measured as the ratio of corporate business (nominal) net value added to corporate business (nominal) net fixed assets at replacement cost. Net value added of corporate business is obtained from BEA NIPA table 1.14, and is available as a quarterly series between 1948:I and 2014:IV. Net fixed assets of corporate business are available as an annual series of year–end estimates in BEA NIPA table 6.1. Using the 1947 year–end estimate as a starting point, these observations are interpolated to generate a quarterly series. The real GDP to CBO's potential GDP, see Figure A.1 second column, uses the CBO's estimates of the potential GDP.

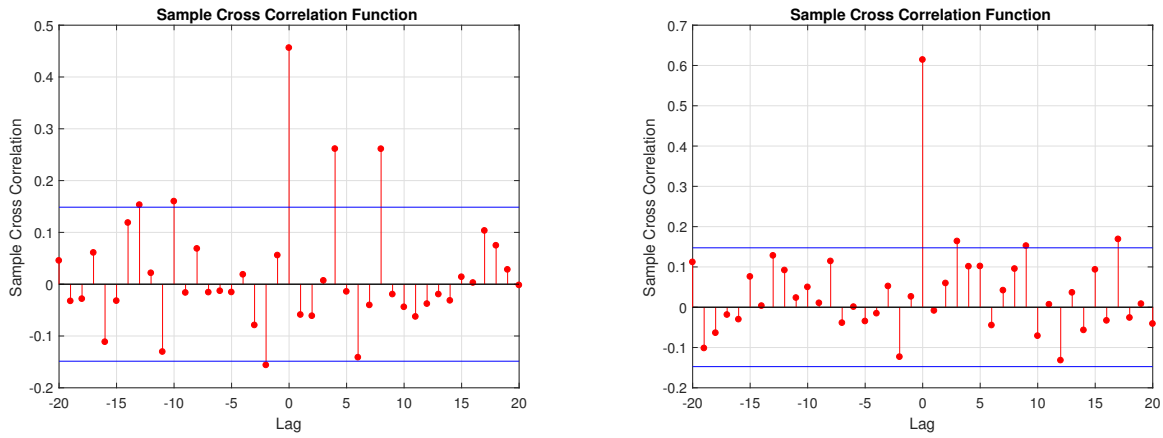
Figure A.1 presents the time series for the income–capital ratio (first column) and real output to CBO's potential output (second column). Their trend is defined as periods larger than 64 quarters per cycle. For the estimation, both time series are normalized between -1 and 1. The NAR estimation results are in Table A.1. The resulting residual analysis and LE bootstrap are in Figures A.2 and A.3, respectively.

Furthermore, Figure A.4 shows the behavior for the LLE with $\omega = 25$. As with the output gap, LLE tends to decrease during peaks and troughs.



(a) Income–Capital ratio (b) GDP to CBO potential putput

Figure A.1: **Income–Capital ratio and GDP to CBO potential output:** These figures exhibit the income–capital ratio (first column) and GDP to CBO's potential GDP (second column). Top panels show the original time series with their trend and the bottom panels the detrended time series used in the estimation.

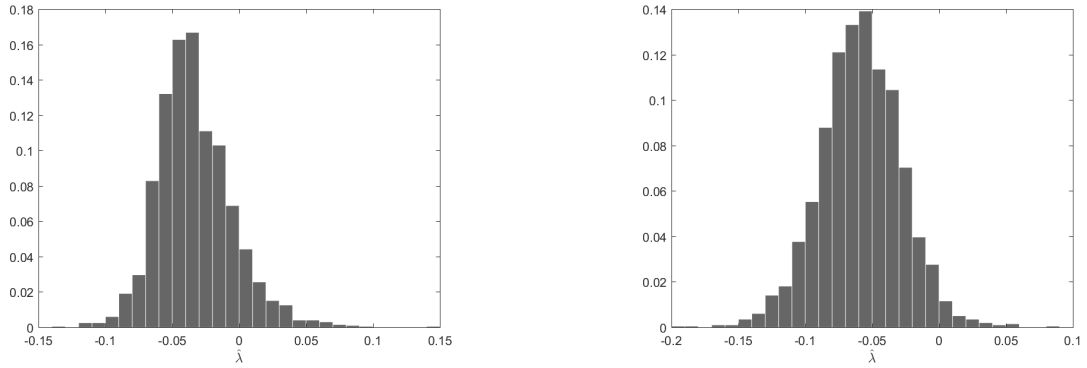


(a) Income–Capital ratio (b) GDP to CBO potential output

Figure A.2: **Analysis of residuals:** These figures show the correlation analysis in equation (2.13) for income–capital ratio (left panel) and real GDP to CBO's potential GDP (right panel).

Table A.1: **Estimation results**

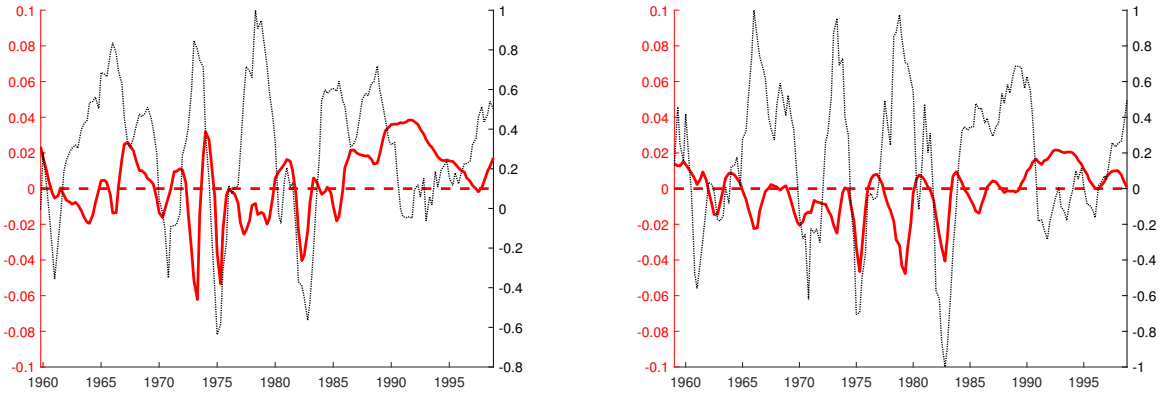
	Income–Capital Ratio		GDP to CBO Potential Output	
		Hidden units (j)		Hidden units (j)
	0	1	0	1
β_j	0.0939	1.3142	0.0741	-1.4562
μ_j		-0.0624		0.0265
Lag (i)		γ_{ji}		γ_{ji}
1		1.0138		-0.7967
2		-0.1193		0.0444
3		-0.0668		0.1137
4		-0.1508		-0.0212
5		-0.0139		0.1067
6		0.1072		-0.0876
7		0.0050		0.0604
8		-0.2758		0.0938
9		0.2174		-0.1498
10		0.0642		-0.0113
11		0.0521		0.0133
12		-0.2686		0.0918
13		0.1693		–
14		-0.0756		–
15		0.0265		–
Root MSE		0.1020		0.1477
GCV(1.4)		0.0121		0.0246
R^2		95.15%		92.86%
Dominant Lyapunov exponent		-0.0547		-0.0409
CI(10%)		(-0.0730,0.0167)		(-0.1082,-0.0091)



(a) Income–Capital ratio

(b) GDP to CBO potential output

Figure A.3: **Dominant Lyapunov exponents bootstrap distribution:** These figures illustrate the distribution of the dominant Lyaunov exponent, which is used to calculate the confidence interval, at 10% of significance, in Table A.1, respectively. See Appendix B.3 for further details on bootstrapping.



(a) Income–Capital ratio

(b) GDP to CBO potential output

Figure A.4: **Local Lyapunov exponent (LLE):** These figures present the local Lyapunov exponent (LLE) for $\omega = 25$ quarters (see subsection 2.3.1 for further details). Left panel for income–capital ratio and right panel real GDP to CBO's potential GDP.

APPENDIX B

CHAPTERS 2 AND 3

B.1 Maximal overlap discrete wavelet transform (MODWT)

B.1.1 Discrete wavelet transform (DWT)

The description of the MODWT in this appendix follows Percival and Walden (2000) and Percival and Mofjeld (1997). Before the description of MODWT, we will start describing the Discrete Wavelet Transform (DWT). The DWT of level J of the N dimensional time series vector $\boldsymbol{X}$ is given by $\boldsymbol{W} = \mathcal{W}\boldsymbol{X}$, where $\boldsymbol{W}$ is a N dimensional vector and $\mathcal{W}$ is a $N \times N$ matrix, satisfying $\mathcal{W}^T\mathcal{W} = I$. The first $N - N/2^J$ contains the wavelet coefficients and the rest $N/2^J$ the scaling coefficients. Therefore, $\boldsymbol{W}$ can be partitioned as

$$\boldsymbol{W} = [\boldsymbol{W}_1^T \boldsymbol{W}_2^T \dots \boldsymbol{W}_J^T \boldsymbol{V}_J^T]$$

where T is the transpose, $\boldsymbol{W}_i$ are of dimensions $N/2^i, i = 1, \dots, J$ and $\boldsymbol{V}_J$ of the remaining $N/2^J$. Matrix $\mathcal{W}$ is built using the wavelet filter $h_{1,0}, \dots, h_{1,L_1-1}$, where L_1 is the width of the first level wavelet filter, with the following properties:

$$\sum_{l=0}^{L_1-1} h_{1,l} = 0, \tag{B.1}$$

$$\sum_{l=0}^{L_1-1} h_{1,l}^2 = 1, \tag{B.2}$$

$$\sum_{l=0}^{L_1-1} h_{1,l} h_{1,l+2n} = \sum_{l=-\infty}^{\infty} h_{1,l} h_{1,l+2n} = 0. \tag{B.3}$$

The filter $h_{1,n}$ has a band–pass given by the interval of frequencies $[1/4, 1/2]$. To define the wavelet filter $h_{j,n}$ for higher scales j, let

$$H_{1,k} = \sum_{n=0}^{N-1} h_{1,n} e^{-i2\pi nk/N}, \quad k = 0, \dots, N-1, \tag{B.4}$$

be the Discrete Fourier Transform (DFT) of the wavelet filter padded with $N - L_1$ zeros. The analysis wavelet, $h_{[1,\cdot]}$, and scaling, $g_{[1,\cdot]}$, filters and synthesis wavelet, $\bar{h}_{[1,\cdot]}$,

and scaling, $\bar{g}_{[1,\cdot]}$, filters are interrelated by $g_{1,j} = (-1)^j h_{1,L-j-1}$, $\bar{h}_{1,j} = h_{1,L-j-1}$, and $\bar{g}_{1,j} = (-1)^{j+1} h_{1,j}$. Let $G_{1,k}$ the DFT of zero–padded scaling filter. The higher–order wavelet filters are defined as

$$h_{j,k} \equiv \frac{1}{N} \sum_{k=0}^{N-1} H_{j,k} e^{i2\pi nk/N}, \tag{B.5}$$

where

$$H_{j,k} \equiv H_{1,2^{j-1}k \bmod N} \prod_{l=0}^{j-2} G_{1,2^l k \bmod N}. \tag{B.6}$$

Elements $h_{j,L_j}, h_{j,L_j+1}, \ldots, h_{j,N-1}$ are equal to zero when $L_j \equiv (2^j - 1)(L_1 - 1) + 1 < N$. The filter $\boldsymbol{h}_j = [h_{j,0}, h_{j,1}, \ldots, h_{j,N-2}, h_{j,N-1}]^T$ is a band–pass filter with approximate frequency band $[1/2^{j+1}, 1/2^j]$. The Jth scaling filter $g_{J,k}$ is defined as

$$g_{J,n} \equiv \frac{1}{N} \sum_{k=0}^{N-1} G_{J,k} e^{i2\pi nk/N}, \tag{B.7}$$

where

$$G_{J,k} \equiv \prod_{l=0}^{J-1} G_{1,2^l k \bmod N}. \tag{B.8}$$

The filter $\boldsymbol{g}_J = [g_{J,0}, g_{J,1}, \ldots, g_{J,N-2}, g_{J,N-1}]^T$ is associated with a low–pass filter $[0, 1/2^{J+1}]$. The first $N - N/2^J$ rows of the matrix $\mathcal{W}$ contain the circularly shifted wavelet coefficients, for instance, the first $N/2$ columns in $\mathcal{W}^T$ are $\mathcal{T}^{2k-1}\boldsymbol{h}_1, k = 1, \ldots, N/2$ where $\mathcal{T}$ is a $N \times N$ matrix that circularly shifts $\boldsymbol{h}_j$ by one unit. The last $N/2^J$ columns of $\mathcal{W}^T$ contain shifted versions of the Jth order scaling filter $\boldsymbol{g}_J$, i.e., $\mathcal{T}^{2^J k-1}\boldsymbol{g}_J, k = 1, \ldots, N/2^J$.

Since $\mathcal{W}$ is orthonormal the original time series, $\boldsymbol{X}$, can be reconstructed by $\boldsymbol{X} = \mathcal{W}^T\boldsymbol{W}$. Furthermore, this equation can be decomposed in a multiresolution fashion, i.e., $\boldsymbol{X}$ can be reconstructed from a series of time series related to variation of the original time series at certain scales. To do this, vector $\boldsymbol{W}$ is partitioned as before,

$$\boldsymbol{W} = [\boldsymbol{W}_1^T, \boldsymbol{W}_2^T, \ldots, \boldsymbol{W}_J^T, \boldsymbol{V}_J^T] \tag{B.9}$$

and the columns of $\mathcal{W}^T$ are partitioned according to the partitioning of $\boldsymbol{W}$, i.e.,

$$\mathcal{W}^T = [\mathcal{W}_1 \mathcal{W}_2 \ldots \mathcal{W}_J \mathcal{V}_J] \tag{B.10}$$

where $\mathcal{W}_j$ is an $N \times N/2^j$ matrix and $\mathcal{V}_J$ is an $N \times N/2^J$ matrix. Hence, the multiresolution analysis (MRA) is

$$\boldsymbol{X} = \mathcal{W}^T\boldsymbol{W} = \sum_{j=1}^{J} \mathcal{W}_j^T\boldsymbol{W}_j + \mathcal{V}_J^T\boldsymbol{V}_J \equiv \sum_{j=1}^{J} \boldsymbol{\mathcal{D}}_j + \boldsymbol{\mathcal{S}}_J \tag{B.11}$$

where $\boldsymbol{\mathcal{D}}_j$ and $\boldsymbol{\mathcal{V}}_J$ are the jth order *detail* and Jth order *smooth*, respectively.

B.1.2 Maximal overlap DWT

The MODWT of level J for time series $\boldsymbol{X}$ defines a highly redundant and nonorthogonal transform with column vectors $\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \ldots, \tilde{\boldsymbol{W}}_J, \tilde{\boldsymbol{V}}_J$ each with dimension N. The MODWT wavelet and scaling coefficients are rescaled versions of the DWT analogous, i.e., $\tilde{\boldsymbol{h}}_j = \boldsymbol{h}_j/2^{j/2}, j = 1, \ldots, J$, and $\tilde{\boldsymbol{g}}_J = \boldsymbol{g}_J/2^{J/2}$. The elements of $\tilde{W}_{j,k}$ are given by $\boldsymbol{X}^T\mathcal{T}^k\tilde{\boldsymbol{h}}_j, k = 0, \ldots N-1$, and the elements of $\tilde{\boldsymbol{V}}_J$ are given by $\boldsymbol{X}^T\mathcal{T}^k\tilde{\boldsymbol{g}}_J$. The DFT for the wavelet, equation (B.6), and scaling, equation (B.8), filter at level j are now $\tilde{H}_{j,k} = H_{j,k}/2^j, j = 1, \ldots, J$ and $\tilde{G}_{J,k} = G_{j,k}/2^J$, respectively. As for the DWT, the MODWT yields also a MRA given by

$$\boldsymbol{X} = \sum_{j=1}^{J} \mathcal{D}_j + \mathcal{V}_J. \tag{B.12}$$

Specifically, I use LA(8) with 4 vanishing moments, $L_1 = 8$, and $J = 5$. Figures B.1 presents synthesis and analysis wavelet and scaling coefficients for level $j = 1$ and Figure B.2 shows the squared gain function showing the frequency response of each detail level and the smooth, which is used as a trend.

B.2 Continuous wavelet transform

The Continuous Wavelet Transform (CWT) maps a time series from time domain into timescale space.[1]

Given the time series $x(t) \in \mathrm{L}^2(\mathbb{R})$, its CWT regarding the mother wavelet $\psi(t)$ is defined as an inner product of $x(t)$ with the family $\psi_{\tau,s}(t)$ of wavelet daughters:

$$W_{x;\psi}(\tau, s) = \langle x(t), \psi_{\tau,s}(t) \rangle = \int_{-\infty}^{+\infty} x(t)\psi^*_{\tau,s}(t)dt, \tag{B.13}$$

with $(^*)$ as the complex conjugate and daughter wavelet function as $\psi_{\tau,s} = |s|^{-1/2}\psi((t-\tau)/s), \tau, s \in \mathbb{R}, s \neq 0$. The position of the mother wavelet function in both time and scale is governed by two parameters s and τ, i.e., the scaling and the translation parameters, respectively.

The mother wavelet must fulfill two requirements, first $\psi(t) \in \mathrm{L}^2(\mathbb{R})$ and second the so–called admissibility condition (a positive constant):

$$0 < C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|}{|\omega|} d\omega < +\infty \tag{B.14}$$

[1] In the case of the Morlet wavelet, one can use exchangeably scale and frequency; however, this relationship is not always that easy to establish. The Morlet wavelet is defined as $\psi(t) = \pi^{-1/4}e^{i\omega_0 t}e^{t^2/2}$.

where $\Psi(\omega)$ is the Fourier transform of the mother wavelet and ω is the angular frequency.

The importance of the admissibility condition is that it allows us to perfectly recover the original time series from its CWT, i.e.,

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} W_{x;\psi}(\tau, s) \psi_{\tau,s} d\tau \right] \frac{ds}{s^2}, s \neq 0, \tag{B.15}$$

and it preserves energy, i.e.,

$$||x||^2 = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} \left| W_{x;\psi}(\tau, s) \right| d\tau \right] \frac{ds}{s^2} \tag{B.16}$$

The power spectrum can be interpreted as the *local* variance for $x(t)$, which is given by

$$\text{WPS}_x(\tau, s) = |W_{x;\psi}(\tau, s)|^2. \tag{B.17}$$

A second magnitude that resembles the Fourier spectral density is the global power spectrum, which averages the spectrum along time,

$$\text{GWPS}_x(s) = \int_{-\infty}^{\infty} \text{WPS}_x(\tau, s) d\tau. \tag{B.18}$$

B.3 Bootstrap for the dominant Lyapunov exponent

To calculate the empirical distribution of the dominant Lyapunov exponent for all the models, I use the 'resampling error' approach (Beninca et al., 2008; Davison & Hinkley, 1997). Let $\hat{F}(\cdot)$ be the function estimated by the feedfordward neural network. From this, I obtain the one–step–ahead forecasting error, defined as $e_t = x_t - \hat{F}(X_{t-1})$. Then each bootstrap sample is generated by first resampling with replacement from $\{e_t\}_{t=1}^N$ to generate the series $\{e_t^*\}_{t=1}^N$. With the resampled errors, I create fictitious 'one–step–ahead' data

$$x_t^* = \hat{F}(X_{t-1}) + e_t^*. \tag{B.19}$$

For each data set, I refitted the neural network model with x^* as the new outputs (or responses) and X_{t-1}, the real data, as inputs (or predictors). I used GCV(1.4) to estimate the number of hidden layers. I allow the model to choose between 1 to 3 hidden layers with 2,000 bootstrap replications.

B.4 Computation of LE and LLE using the QR–factorization

In order to calculate the approximation of LE and LLE, we need to use the QR–factorization from the matrix product in equation (2.3) (see von Bremen, Udwadia, and Proskurowski (1997)). We rewrite equation (2.3) the n matrix multiplication as $J_n J_{n-1} \dots J_1$.

This decomposition can be done sequencially, i.e., starting from $Q_0 = I$,

$$\begin{aligned}
\mathrm{qr}[J_n J_{n-1} \dots J_1] &= \mathrm{qr}[J_n J_{n-1} \dots J_2 (J_1 Q_0)] = \mathrm{qr}[J_n J_{n-1} \dots (J_2 Q_2)][R_1] \\
&= \mathrm{qr}[J_n J_{n-1} \dots (J_3 Q_2)][R_2 R_1] = \dots \\
&= \mathrm{qr}[J_n J_{n-1} \dots (J_i Q_{i-1})][R_{i-1} R_{i-2} \dots R_2 R_1] = \dots \\
&= Q_n [R_n R_{n-1} \dots R_2 R_1] = Q_n R.
\end{aligned}$$

Each of the diagonals terms in R are the multiplication of R_i's diagonal elements. Therefore, the m LE and LLE in equations (2.4) and (2.6) can be calculated as $\lambda_k = n^{-1} \sum_{j=1}^{n} \ln |R_i(k,k)|, k = 1, \dots, n$.

B.5 Levenberg–Marquardt algorithm

Levenberg–Marquardt learning algorithm is a combination of the Steepest Decent and Gauss–Newton methods. Given a quadratic performance index $F(\boldsymbol{\omega})$, the Gauss–Newton's method is

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \boldsymbol{A}_k^{-1} \boldsymbol{g}_k, \tag{B.20}$$

where $\boldsymbol{\omega} \in \mathbb{R}^m$, $\boldsymbol{A}_k \equiv \nabla^2 F(\boldsymbol{\omega})\big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_k} \in \mathbb{R}^{m \times m}$, and $\boldsymbol{g}_k \equiv \nabla F(\boldsymbol{\omega})\big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_k} \in \mathbb{R}^m$.

Given the performance index, $F(\boldsymbol{\omega}) = \boldsymbol{\varepsilon}(\boldsymbol{\omega})^T \boldsymbol{\varepsilon}(\boldsymbol{\omega})$ and its gradient and Hessian might be written as follows:

$$\nabla F(\boldsymbol{\omega}) = 2\boldsymbol{J}(\boldsymbol{\omega})^T \boldsymbol{\varepsilon}(\boldsymbol{\omega}) \quad \text{and} \quad \nabla^2 F(\boldsymbol{\omega}) = 2\boldsymbol{J}(\boldsymbol{\omega})^T \boldsymbol{J}(\boldsymbol{\omega}) + 2\boldsymbol{S}(\boldsymbol{\omega})$$

where

$$\boldsymbol{J}(\boldsymbol{\omega}) = \begin{bmatrix} \frac{\partial \varepsilon_1(\boldsymbol{\omega})}{\partial \omega_1} & \frac{\partial \varepsilon_1(\boldsymbol{\omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_1(\boldsymbol{\omega})}{\partial \omega_m} \\ \frac{\partial \varepsilon_2(\boldsymbol{\omega})}{\partial \omega_1} & \frac{\partial \varepsilon_2(\boldsymbol{\omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_2(\boldsymbol{\omega})}{\partial \omega_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \varepsilon_n(\boldsymbol{\omega})}{\partial \omega_1} & \frac{\partial \varepsilon_n(\boldsymbol{\omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_n(\boldsymbol{\omega})}{\partial \omega_m} \end{bmatrix} \tag{B.21}$$

and

$$\boldsymbol{S}(\boldsymbol{\omega}) = \sum_{i=1}^{n} \varepsilon_i(\boldsymbol{\omega}) \nabla^2 \varepsilon_i(\boldsymbol{\omega})$$

If we assume that $\boldsymbol{S}(\boldsymbol{\omega})$ is small, the Hessian becomes $\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J}$. Therefore, Gauss–Newton's method can be written as:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \left[\boldsymbol{J}(\boldsymbol{\omega}_k)^T\boldsymbol{J}(\boldsymbol{\omega}_k)\right]^{-1}\boldsymbol{J}(\boldsymbol{\omega}_k)^T\boldsymbol{\varepsilon}(\boldsymbol{\omega}_k), \tag{B.22}$$

The Levenberg–Marquardt algorithm modifies the above equation in the following way:

$$\boldsymbol{\omega}_{k+1} = \boldsymbol{\omega}_k + \left[\boldsymbol{J}(\boldsymbol{\omega}_k)^T\boldsymbol{J}(\boldsymbol{\omega}_k) + \mu_k\boldsymbol{I}_m\right]^{-1}\boldsymbol{J}(\boldsymbol{\omega}_k)^T\boldsymbol{\varepsilon}(\boldsymbol{\omega}_k), \tag{B.23}$$

If μ_k is close to zero, equation (B.23) becomes the Gauss–Network method. However, if μ_k is very large, it becomes into the Steepest Decent algorithm.

B.6 Bayesian regularization

Generalization (or fear of over or under fitting) is a major concern in ANN. Given the use of the NAR model, the user choices are the number of delays, squashing function, and number of hidden units, i.e., the complexity of the model. Broadly speaking, there are two ways to constrain the complexity a model: first, by restricting the number of parameters and, second, by constraining their growth.

In this paper, I use the latter through the Bayesian regularization, which was proposed by MacKay (1992) and was first applied to backpropagation by Foresee and Hagan (1997). Under regularization, the performance index becomes:

$$F(\boldsymbol{\omega}) = \beta E_D(D|\boldsymbol{\omega}, M) + \alpha E_W(\boldsymbol{\omega}|M) \tag{B.24}$$

where E_D is the MSE defined in (3.10), D is the training set with input–output pairs, M is the network architecture, $E_W(\boldsymbol{\omega}|D) = \sum_{i=1}^{m}\omega_j^2$, and α, β are hyperparameters that need to be estimated.

The posterior distribution of the ANN weights can be updated according the Bayes' rule:

$$P(\boldsymbol{\omega}|D, \alpha, \beta, M) = \frac{P(D|\boldsymbol{\omega}, \beta, M)P(\boldsymbol{\omega}|\alpha, M)}{P(D|\alpha, \beta, M)} \tag{B.25}$$

with $P(\boldsymbol{\omega}|\alpha, M)$ the prior distribution of weights, $P(D|\boldsymbol{\omega}, \beta, M)$ is the likelihood function, which is the probability density of the data given the network weights, and $P(D|\alpha, \beta, M)$ is called evidence or the normalization factor. Assuming that the noise of the training set

and the distribution of the weights and biases are Gaussian, the probability densities can be written as

$$P(\boldsymbol{\omega}|\alpha, M) = \frac{1}{Z_W(\alpha)} \exp\left(-\alpha E_W\right),$$

and

$$P(D|\beta, M) = \frac{1}{Z_D(\beta)} \exp\left(-\beta E_D\right),$$

with $Z_W(\alpha) = (\pi/\alpha)^{m/2}$, $Z_D(\beta) = (\pi/\beta)^{T/2}$, $\alpha = 1/(2\sigma_\varepsilon^2)$, $\beta = 1/(2\sigma_\omega^2)$, m number of network parameters, T number of observations, and σ_ω^2 and σ_ε^2 are the variance of the network weights and training error, respectively.

Then, the posterior probability can be written as

$$\begin{aligned} P(\boldsymbol{\omega}|D, \alpha, \beta, M) &= \frac{\frac{1}{Z_W(\alpha)} \frac{1}{Z_D(\beta)} \exp\left(-(\beta E_D + \alpha E_W)\right)}{\text{Normalizing Factor}} \\ &= \frac{1}{Z_F(\alpha, \beta)} \exp\left(-F(\boldsymbol{\omega})\right) \end{aligned} \tag{B.26}$$

Clearly, maximizing the posterior probability of $\boldsymbol{\omega}$ is equivalent to minimizing the regularized performance index equation (B.24).

Since α, β are assumed to be parameters in the Bayesian framework, their posterior can be calculated as

$$P(\alpha, \beta|D, M) = \frac{P(D|\alpha, \beta, M) P(\alpha, \beta|M)}{P(D|M)} \tag{B.27}$$

It is further assumed that $P(\alpha, \beta|M)$ is uniform; therefore, maximizing the posterior distribution translates into maximizing the likelihood function $P(D|\alpha, \beta, M)$, which is normalization factor in equation (B.25), hence

$$\begin{aligned} P(D|\alpha, \beta, M) &= \frac{P(D|\boldsymbol{\omega}, \beta, M) P(\boldsymbol{\omega}|\alpha, M)}{P(\boldsymbol{\omega}|D, \alpha, \beta, M)} \\ &= \frac{\frac{1}{Z_W(\alpha)} \exp\left(-\alpha E_W\right) \frac{1}{Z_D(\beta)} \exp\left(-\beta E_D\right)}{\frac{1}{Z_F(\alpha,\beta)} \exp\left(-F(\boldsymbol{\omega})\right)} \\ &= \frac{Z_F(\alpha, \beta)}{Z_W(\alpha) Z_D(\beta)} \end{aligned} \tag{B.28}$$

where $Z_W(\alpha)$ and $Z_D(\beta)$ are known.

To calculate Z_F, the performance index can be expanded in a second order Taylor expansion around its local minimum point, $\boldsymbol{\omega}^{MP}$, where the the shape is quadratic.

$$F(\boldsymbol{\omega}) \approx F\left(\boldsymbol{\omega}^{MP}\right) + \frac{1}{2}\left(\boldsymbol{\omega} - \boldsymbol{\omega}^{MP}\right)^T \boldsymbol{H}^{MP}\left(\boldsymbol{\omega} - \boldsymbol{\omega}^{MP}\right), \tag{B.29}$$

where $\boldsymbol{H}$ is the Hessian matrix of the performance index, i.e., $\boldsymbol{H} = \beta\nabla^2 E_D + \alpha\nabla^2 E_W$, and replacing equation (B.29) into (B.26), the posterior distribution can be written as

$$P(\boldsymbol{\omega}|D, \alpha, \beta, M) = \left\{\frac{1}{Z_F}\exp\left(-F\left(\boldsymbol{\omega}^{MP}\right)\right)\right\}\exp\left\{\frac{1}{2}\left(\boldsymbol{\omega} - \boldsymbol{\omega}^{MP}\right)^T \boldsymbol{H}^{MP}\left(\boldsymbol{\omega} - \boldsymbol{\omega}^{MP}\right)\right\}, \tag{B.30}$$

for the above density to be equal to the normal density

$$Z_F(\alpha, \beta) \approx (2\pi)^{m/2}\left(\left|\boldsymbol{H}^{MP}\right|^{-1}\right)^{1/2}\exp\left(-F\left(\boldsymbol{\omega}^{MP}\right)\right). \tag{B.31}$$

Then, placing the above equation into (B.28) and maximizing it yields

$$\alpha^{MP} = \frac{\gamma}{2E_W(\boldsymbol{\omega}^{MP})} \quad \text{and} \quad \beta^{MP} = \frac{m - \gamma}{2E_D(\boldsymbol{\omega}^{MP})}, \tag{B.32}$$

where $\lambda = m - 2\alpha^{MP}\mathrm{Tr}(\boldsymbol{H}^{MP})^{-1}$ is the so–called effective number of parameters, which is between zero and n. The effective number of parameters measures how many parameters effectively reduce the error function.

The steps require for Bayesian optimization of the regularization parameters, with the Gauss–Newton approximation to Hessian, which is readily available if the Levenberg–Marquardt algorithm is used.

0. Initialize α, β and the weights. The weights are initialized randomly. Then E_D and E_W are computed. Set $\lambda = m$, and calculate α and β.

1. Take one step of the Levenberg–Marquardt algorithm toward minimizing the objective function $F(\boldsymbol{\omega}) = \beta E_D + \alpha E_W$.

2. Compute the effective number of parameters, λ, making use of the Gauss–Newton approximation to the Hessian available in the Levenberg–Marquardt algorithm (see Appendix B.5): $\boldsymbol{H} = \nabla^2 F(\boldsymbol{\omega}) \approx 2\beta\boldsymbol{J}^T\boldsymbol{J} + 2\alpha\boldsymbol{I}_n$, where $\boldsymbol{J}$ is the Jacobian matrix of the training set error (see equation (B.21)).

3. Compute the new estimates for the regularization parameters α and β.

4. Iterate steps 1 through 3 until convergence.

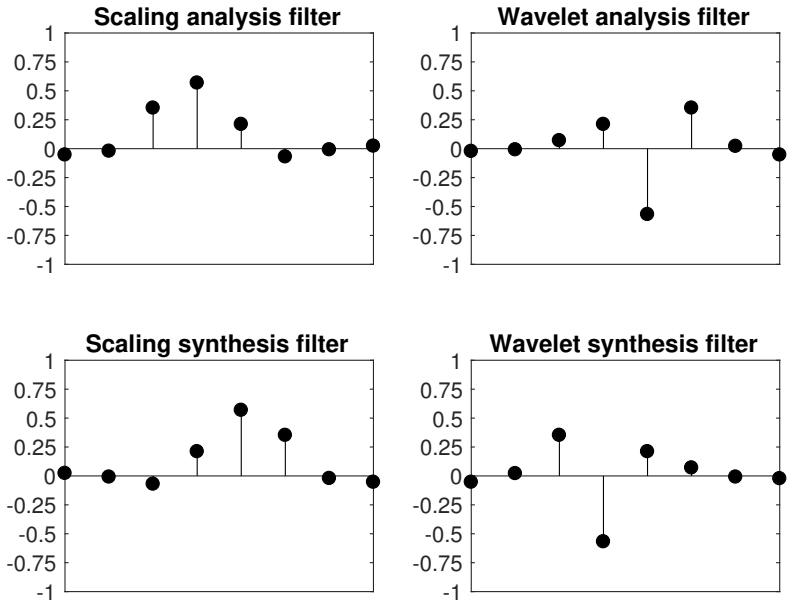


Figure B.1: **MODWT wavelet and scaling filter coefficients:** This figure presents the MODWT synthesis and analysis wavelet and scaling filter coefficient for LA(8).

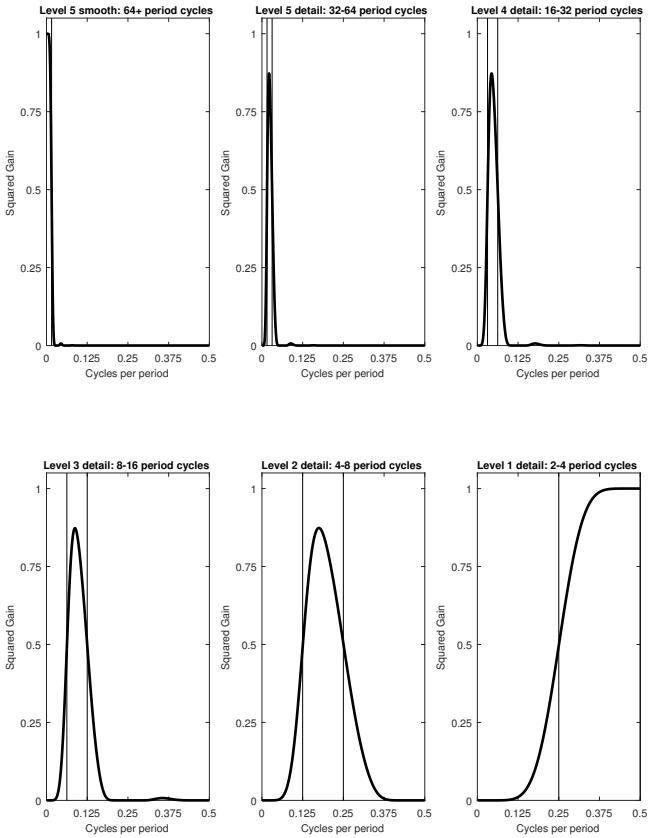


Figure B.2: **Square gain function for MODWT wavelet and scaling filter coefficients:** This figure presents the squared gain function for the synthesis wavelet and scaling filter, Figure B.1, with $J = 5$.