Department of Economics

Athens University of Economics and Business

WORKING PAPER no. 09-2023

A machine learning approach to construct quarterly data on intangible investment for Eurozone

Angelos Alexopoulos and Petros Varthalitis

May 2023

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

The Working Papers in this series circulate mainly for early presentation and discussion, as well as for the information of the Academic Community and all interested in our current research activity.

The authors assume full responsibility for the accuracy of their paper as well as for the opinions expressed therein.

A machine learning approach to construct quarterly data on intangible investment for Eurozone

Angelos Alexopoulos* Petros Varthalitis† ‡

May 26, 2023

Abstract

We develop a novel approach to construct quarterly time series data for annually measured intangible investment variables. We accomplish this by using machine learning methods to explore the relationship between these variables and key macroeconomic time series available on a quarterly frequency. The proposed approach offers some advantages over other econometric techniques. Specifically, it does not require any ex-ante assumptions for the link between the quarterly time series and their annual counterparts, and it is free from issues such as multicollinearity and endogeneity, requiring almost no data pre-processing. To demonstrate the usefulness of the constructed data, we present some business cycles facts for the intangible economies of Eurozone and estimate a dynamic factor model.

Keywords: Machine Learning, Intangible Investment, Factor Model, Business Cycles

JEL Classification: E01, E22, E32, C49, C80

1 Introduction

Intangible assets are increasingly important in modern economies, but they are difficult to measure accurately. Corrado et al. (2016) have developed a comprehensive database that provides harmonized cross-country data on intangible investment for 15 EU countries. Using these data, Bontadini et al. (2023) have extended the widely known EUKLEMS productivity database with intangible investment (see EUKLEMS & IntanProd). However, these data are available only on an annual frequency. On the other hand, most key macroeconomic variables, such as gross domestic product, gross value added (GVA), consumption, and investment, are available on a quarterly frequency. Having harmonized data related to the intangible economy on a quarterly frequency is crucial for policymakers, academics, and practitioners to construct empirically relevant macroeconomic models and make accurate macroeconometric inferences. The need for such data has become more pronounced in recent decades, as the intangible intensity of modern economies has risen rapidly. In Eurozone, the share of intangible investment has steadily increased from 9.5% in 1995 to 14% in 2016, while the ratio of intangible to tangible investment has almost doubled.[1]

Our aim is to construct quarterly time series for the intangible economy of Eurozone countries. While various econometric techniques deal with mixed frequency data their primary objective is to enhance the accuracy of macroeconomic forecasting. In contrast, our focus is distinct in that we prioritize using all the available information of (macroeconomic) indicators observed at lower frequency to construct the quarterly time series.[2]. Here, we adopt a machine learning approach (see e.g., Ballarin et al. (2022)), namely, the extreme gradient boosting (XGboost) regression method (Friedman, 2001; Chen and Guestrin, 2016) to construct quarterly time series. The XGboost regression relies on an ensemble of decision trees to construct a prediction model for a target variable by using a set of auxiliary variables.

Our approach has several merits. First, it does not require any ex-ante assumptions for the link between the high frequency and the low frequency variables of interest. Second, it is free of issues commonly encountered in OLS estimation such as multicollinearity and endogeneity, and it does not require complicated and

*Department of Economics, Athens University of Economics and Business, Greece. Email: angelos@aueb.gr.

†Department of Economics, Athens University of Economics and Business, Greece. Email: pvarthalitis@aueb.gr.

‡Corresponding author

[1]See Figures S.1 and S.2 in the supplementary material. A similar trend is also observed in the US economy, where intangible investment amounted to 17% of GDP in 2021, compared to only 8.5% for tangible investment. For a recent review see Corrado et al. (2022).

[2]These include but are not limited to: (i) the bridge model approach (Baffigi et al., 2004); (ii) the mixed data sampling (MIDAS) regressions (Ghysels et al., 2007); and (iii) vector autoregressive and dynamic factor models (Mariano and Murasawa, 2010); see Foroni et al. (2015) for a review.

computationally expensive estimation algorithms, e.g., Bayesian techniques. Third, it does not rely on assumptions usually made in standard econometric time series (e.g., stationarity) and thus requires almost no data preprocessing. These characteristics make the proposed approach easy to implement by macroeconomists and other practitioners. We demonastrate the performance of the machine learning approach on macroeconomic variables that are available on both annual and quarterly frequency (e.g., GVA). Subsequently, we construct quarterly data for the variables for which we only have annual observations, namely GVA adjusted for intangible assets not measured in national accounts, intangibles assets not measured in national accounts and total intangibles assets.

To illustrate the usefulness of the constructed data we derive some key business cycles statistics for Eurozone intangible economies. Then, we estimate a dynamic factor model to measure the degree of synchronization among these economies.[3] We find that intangible investment is two or three times more volatile as GVA and procyclical; whereas intangible investment is less volatile and less procyclical than tangible investment. Our estimated global factor captures recent trends observed in standard Eurozone macroeconomic aggregates quite well. Interestingly, our analysis implies a relatively high degree of synchronicity among the intangible economies of Eurozone countries of our sample and also provides evidence about the existence of global drivers behind the fluctuations of intangible economies.

Our paper contributes to the existing literature in the following ways. Firstly, we propose a novel approach to construct high-frequency macroeconomic variables using low-frequency macroeconomic variables by adopting a machine learning method. This approach is easy to implement and relaxes some of the strong assumptions used in other popular econometric techniques. Secondly, we construct quarterly time-series measurements of the intangible economy for Eurozone countries. This dataset enables macroeconomists and practitioners to utilize the data for macroeconomic modeling and policy evaluation. Finally, we utilize the constructed quarterly time-series to measure the business cycle characteristics of intangible economies and the degree of their synchronicity in Eurozone countries. The rest of the paper is organized as follows. Section 2 presents the developed data construction method, in Section 3 we present our results and analysis and Section 4 closes the paper. For readers interested in more technical details an online supplementary appendix is available.

2 Methodology

2.1 Data

We employ data from two sources. First, we draw annual macroeconomic aggregates related to intangible economy from the EUKLEMS & INTANPROD database which comprises harmonized time series for intangible investment over the period 1995-2018.[4] Second, we draw standard quarterly macroeconomic aggregates from the national accounts of Eurostat for the same period. Table 1 lists our dataset, panel A presents the time series that are available only in annual frequency, while panel B presents the macroeconomic time series that we use to construct quarterly observations for the variables in panel A.

[3]Estimating such a model would be problematic without the construction of quarterly time series, as annual data only spans for a 23-year period.

[4]The countries in our sample include 14 Eurozone countries, namely, Austria, Belgium, Estonia, Finland, France, Germany, Ireland, Italy, Netherlands, Latvia, Lithuania, Luxembourg, Portugal, Slovenia, Slovakia and Spain while we exclude Croatia, Cyprus and Malta due to data availability.

Table 1: Data

Variable	Annual	Quarterly
Panel A		
Gross Value Added adjusted	y_t^{gva}	$(y_{t_q}^{gva})$
Total Intangibles	y_t^{I}	$(y_{t_q}^{I})$
Non-National Accounts Intangibles	y_t^{NI}	$(y_{t_q}^{NI})$
Panel B		
Gross Value Added	$x_t^{(1)}$	$x_{t_q}^{(1)}$
Gross domestic product	$x_t^{(2)}$	$x_{t_q}^{(2)}$
Final consumption expenditure	$x_t^{(3)}$	$x_{t_q}^{(3)}$
Gross fixed capital formation	$x_t^{(4)}$	$x_{t_q}^{(4)}$
Tangible Investment	$x_t^{(5)}$	$x_{t_q}^{(5)}$
Exports of goods and services	$x_t^{(6)}$	$x_{t_q}^{(6)}$
Imports of goods and services	$x_t^{(7)}$	$x_{t_q}^{(7)}$
Compensation of employees	$x_t^{(8)}$	$x_{t_q}^{(8)}$
Wages and salaries	$x_t^{(9)}$	$x_{t_q}^{(9)}$
Total employment domestic concept	$x_t^{(10)}$	$x_{t_q}^{(10)}$
Taxes on production and imports	$x_t^{(11)}$	$x_{t_q}^{(11)}$
Taxes on products	$x_t^{(12)}$	$x_{t_q}^{(12)}$
Final consumption expenditure of general government	$x_t^{(13)}$	$x_{t_q}^{(13)}$
Household and NPISH final consumption expenditure	$x_t^{(14)}$	$x_{t_q}^{(14)}$
Final consumption, durable goods	$x_t^{(15)}$	$x_{t_q}^{(15)}$
Final consumption, semi-durable goods	$x_t^{(16)}$	$x_{t_q}^{(16)}$
Final consumption, non-durables	$x_t^{(17)}$	$x_{t_q}^{(17)}$
Final consumption, services	$x_t^{(18)}$	$x_{t_q}^{(18)}$
Employees domestic concept	$x_t^{(19)}$	$x_{t_q}^{(19)}$
Self-employed domestic concept	$x_t^{(20)}$	$x_{t_q}^{(20)}$
Employees domestic concept	$x_t^{(21)}$	$x_{t_q}^{(21)}$

Notes: Variables in parentheses are not observed at the frequency indicated by the corresponding column. The two columns on the right hand side present the notation used for the value of each variable at time t. Tangible investment $(x_t^{(5)}, x_{t_q}^{(5)})$ is computed by subtracting intellectual property products (gross) from total fixed assets (gross).

2.2 Data construction by using machine learning

We adopt a machine learning approach to construct quarterly time series for which we have access to their annual counterpart. We utilize two sets of observed time series: the variables of interest at an annual frequency and a set of macroeconomic aggregates observed at both annual and quarterly frequencies We denote by y_t the observation at year t ($t = 1, \ldots, T$) and aim to estimate the variables y_{t_q} such that $y_t = \sum_{q=1}^4 y_{t_q}$, where t_q is the q-th quarter of year t. We also denote by $x_{t_q}^{(d)}$ the value of the d-th ($d = 1, \ldots, D$) quarterly observed time series data (macroeconomic features) in the q-th quarter of year t.

The proposed technique relies on the gradient boosting method where an ensemble of decision trees is employed to predict the values of an annually observed variable in each one of the four quarters of the year. First we set $x_t^{(d)} = \sum_{q=1}^4 x_{t_q}^{(d)}$ to be the annual measurement of the the quarterly observed variable $x_{t_q}^{(d)}$, for each $d = 1, \ldots, D$. By employing the XGboost algorithm we learn a function G, such that $\hat{y}_t = G(x_t^{(1)}, \ldots, x_t^{(d)})$ and the quadratic loss $\sum_{t=1}^T (y_t - \hat{y}_t)^2$ is minimised. The function G encodes all the available information about the relationship between the target y_t and the macroeconomic features $\{x_{t_q}^{(d)}\}_{d=1}^D$. Then, we estimate the target variables, y_{t_q}, by applying the function G on the quarterly version of the macroeconomic features, i.e., we set $\hat{y}_{t_q} = G(x_{t_q}^{(1)}, \ldots, x_{t_q}^{(d)})$. Finally, by noting that the sum of the quarterly measured variables over a year t should be equal to the observed value y_t we further improve the estimation of y_{t_q} by calculating the error $\epsilon_t = y_t - \sum_{q=1}^4 \hat{y}_{t_q}$ and setting $\tilde{y}_{t_q} = \hat{y}_{t_q} + \epsilon_t/4$ if $\epsilon_t > 0$ and $\tilde{y}_{t_q} = \hat{y}_{t_q} - \epsilon_t/4$ otherwise. In the supplementary appendix we provide all the technical details of the XGboost algorithm that we used to construct the function

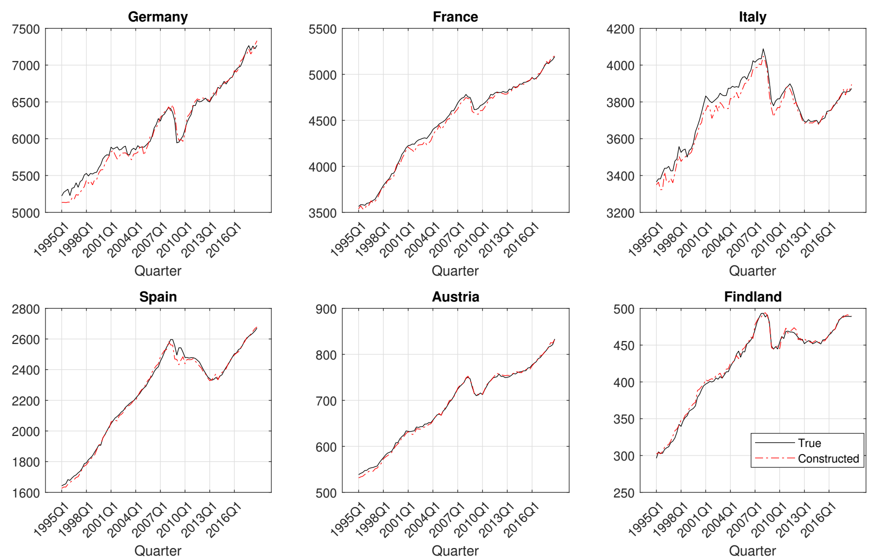
G as well an algorithmic description of our developed method for the calculation of $\tilde{y}_{t_q}$.

3 Results

3.1 Evaluation of the machine learning approach

To evaluate the proposed approach for quarterly data construction we work with a macroeconomic variable for which we have available time series at both annual and quarterly frequencies such as GVA, $x_t^{(1)}$ and $x_{t_q}^{(1)}$ respectively in Table 1. We construct $\tilde{x}_{t_q}^{(1)}$ and we compare it with the true variable $x_{t_q}^{(1)}$. Figure 1 presents the true quarterly time series of gross value added (shown in black solid lines) and the constructed quarterly time series (shown in red dashed lines) for the largest Eurozone economies in our sample. It is clear that our technique performs quite well in replicating both the macroeconomic trend and the cyclical properties of the true quarterly time series. We report that the mean absolute error across the different countries is quite small and ranges between 0.5% and 1% of the country's GVA.

Figure 1: Constructed vs true GVA



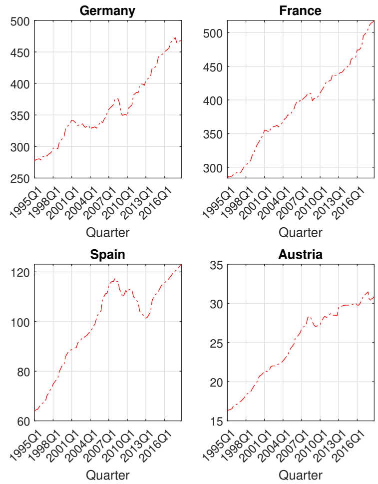
Notes: GVA is scaled with GVA deflator.

3.2 Business cycles facts for the intangible economies of Eurozone using the constructed quarterly time series

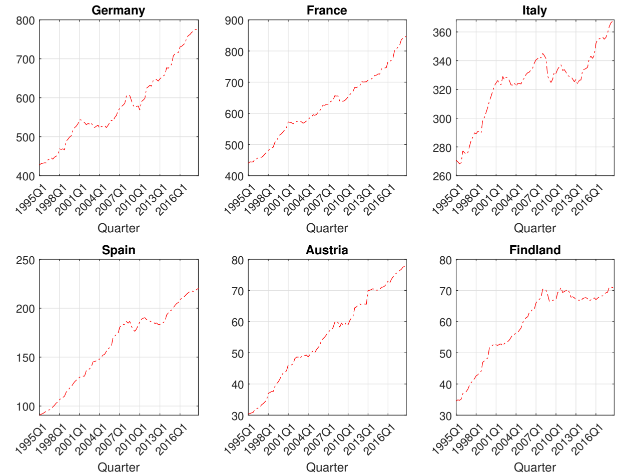
Figure 2 presents the constructed quarterly counterparts for time series that are only available on an annual frequency, namely intangible investment not currently measured in the System of National Accounts (SNA) (a), total intangible investment (b), and GVA adjusted (c) for the six largest economies in the Eurozone.

Figure 2: Constructed quarterly time series for intangible economies

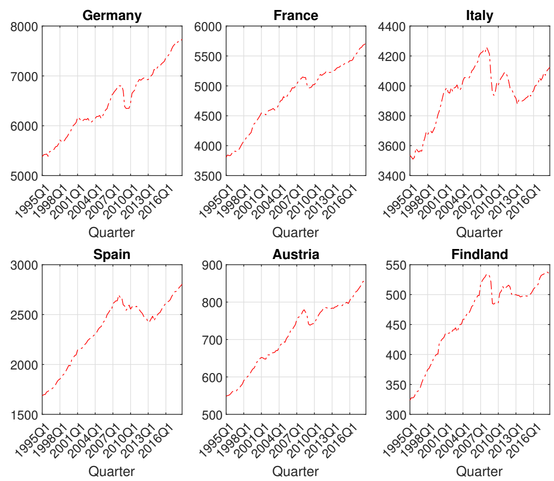
(a) Non-National Accounts Intangible Investment, $\tilde{y}^{NI}_{t_q}$



(b) Total Intangible Investment, $\tilde{y}^{I}_{t_q}$



(c) GVA adjusted, $\tilde{y}^{gva}_{t_q}$



We use these time series to derive the cyclical properties of the Eurozone's intangible economies drawing from the conventional business cycle analysis (see, e.g., Fiorito and Kollintzas (1994)). Table 2 presents the volatility of the constructed series relative to GVA in columns 1 to 3 and to tangible investment in column 4 for each country. Table 3 presents the correlations of the constructed series with GVA in columns 1-3 and the correlation of total intangible to tangible investment in column 4 Tables 2 and 3 reveal some interesting stylized facts about the business cycle properties of the intangible economy in Eurozone countries. Firstly, total intangible investment and intangible investment not measured in the SNA are two to three times more volatile than GVA (see columns 1 and 2 in Table 3). Secondly,intangible investment is less volatile than tangible investment for most countries (see column 4); while GVA adjusted (i.e., GVA augmented with intangible investment) is more volatile than GVA currently reported in the SNA (see column 3). From Table 3 it is clear that intangible investment is positively correlated with GVA (i.e., procyclical) and tangible investment (see columns 1, 2 and 4). The intangible-adjusted GVA correlation with the SNA, GVA is weaker than expected, which makes the accurate measurement of non-national accounts intangibles quite important.

Table 2: Relative volatilities of intangible variables
with respect to GVA and tangible investment

Countries	$\frac{\sigma(\tilde{y}^{NI}_{t_q})}{\sigma(x^{(1)}_{t_q})}$	$\frac{\sigma(\tilde{y}^{I}_{t_q})}{\sigma(x^{(gva)}_{t_q})}$	$\frac{\sigma(\tilde{y}^{gva}_{t_q})}{\sigma(x^{(1)}_{t_q})}$	$\frac{\sigma(\tilde{y}^{I}_{t_q})}{\sigma(x^{(5)}_{t_q})}$
Austria	1.64	2.20	1.06	0.71
Belgium	3.88	3.25	1.16	0.65
Germany	1.37	1.32	0.94	0.48
Estonia	1.35	1.40	1.04	0.49
Spain	1.61	1.52	1.20	0.43
Finland	1.01	1.18	1.01	0.54
France	1.62	1.84	1.20	0.65
Ireland	2.27	3.77	1.41	1.39
Italy	1.20	1.25	1.06	0.41
Lithuania	3.79	3.40	1.45	0.88
Luxembourg	1.00	1.11	0.85	0.24
Latvia	1.98	1.82	1.11	0.46
Netherlands	1.84	5.67	1.41	1.30
Portugal	2.09	2.07	1.43	0.43
Slovenia	1.47	1.47	1.37	0.36
Slovakia	1.57	1.99	0.95	0.51

Notes: We denote by $\sigma(\cdot)$ the volatility of the variable in the parenthesis; $x^{(1)}_{t_q}$ and $x^{(5)}_{t_q}$ denote GVA and tangible investment respectively both observed on quarterly frequency. The volatility of each time series has been computed for the first differences of the log-transformed variables.

Table 3: Correlations of intangible variables with respect to GVA and tangible investment

Countries	$\mathrm{Cor}(\tilde{y}^{NI}_{t_q}, x^{(2)}_{t_q})$	$\mathrm{Cor}(\tilde{y}^{I}_{t_q}, x^{(2)}_{t_q})$	$\mathrm{Cor}(\tilde{y}^{gva}_{t_q}, x^{(2)}_{t_q})$	$\mathrm{Cor}(\tilde{y}^{I}_{t_q}, x^{(21)}_{t_q})$
Austria	0.46	0.36	0.77	0.25
Belgium	0.23	0.21	0.63	0.16
Germany	0.64	0.53	0.85	0.61
Estonia	0.59	0.52	0.75	0.40
Spain	0.46	0.35	0.61	0.58
Finland	0.60	0.56	0.76	0.47
France	0.54	0.54	0.79	0.48
Ireland	-0.29	0.20	0.33	0.23
Italy	0.64	0.59	0.79	0.44
Lithuania	0.37	0.40	0.77	0.23
Luxembourg	0.27	0.30	0.30	0.28
Latvia	0.54	0.54	0.70	0.40
Netherlands	0.21	0.10	0.47	0.12
Portugal	0.50	0.50	0.56	0.45
Slovenia	0.66	0.56	0.69	0.46
Slovakia	0.46	0.45	0.66	0.24

Notes: $x^{(1)}_{t_q}$ and $x^{(5)}_{t_q}$ denote GVA and tangible investment respectively; both observed on quarterly frequency. The correlations have been computed for the first differences of the log-transformed variables in each of the time series.

3.3 Econometric application: A dynamic factor model

We utilize the constructed quarterly data to estimate a standard dynamic factor model. The purpose of this experiment is to illustrate the need for measuring the intangible economy at a quarterly frequency, rather than to provide a meticulous econometric analysis of the intangible economies in the Eurozone. We consider an one-factor model in which the constructed quarterly variables of the i-th country, i.e. the total intangible investment, the non-national account intangible investment and the GVA adjusted are jointly modelled. We set $\tilde{y}_{i,j,t_q}$ to be the measurement constructed for the j-th variable[5], $j = 1, 2, 3$, of the i-th country, $i = 1, \ldots, N$, in the q-th quarter of year t, $t = 1, \ldots, T$ and we assume that,

$$\tilde{y}_{i,j,t_q} = w_{ij} f_{t_q} + e_{i,j,t_q}, \;\; e_{i,j,t_q} \sim N(0, \sigma^2_{ij}), \tag{1}$$

where f_{t_q} denotes the value of the common factor in the q-th quarter of the t year and w_{ij} is the factor loading that corresponds to the j-th variable and the i-th country. To model the time evolution of the latent factor we assume an autoregressive process of order one,

$$f_{t_q} = \phi f_{t_q^-} + \eta_{t_q}, \;\; \eta_{t_q} \sim N(0, \sigma^2_\eta)$$

where $f_{t_q^-}$ denotes the value of the factor one quarter before t_q and $\phi \in (-1, 1)$ to ensure a stationary process. To estimate the model we follow the Bayesian paradigm as Kose et al. (2012) and Berger et al. (2021). After assigning prior distributions to the parameters of the model we employ a Markov Chain Monte Carlo algorithm to draw samples from the joint posterior of the latent factor path, the factor loadings and the variances of the error terms in (1). In the supplementary material we provide all the details for the prior distributions that we assign to the parameters of the model.

Figure 3 illustrates that the estimated global factor captures the macroeconomic cyclical fluctuations (booms/busts) observed in Eurozone the 1995Q1-2018Q4 period. The light grey shaded area depicts the Euphoria period that preceded and followed the introduction of the common currency. The grey shaded area depicts the eruption of the Global Financial Crisis (GFC) in 2007Q1-2009Q4, while the light dark grey illustrates the subsequent European Debt Crisis (EDC), i.e., 2010Q1-2014Q1. The dark grey shaded area shows the period of recovery, i.e., 2014Q2-2018Q4. Clearly, the constructed time series data series exhibit cyclical properties similar to those observed in standard quarterly macroeconomic series of Eurozone.

We also decompose the volatility of the intangible economy in each country to the volatility generated by the common factor and other unobserved characteristics. By inspecting Table 4 we conclude that the global factor accounts for a fraction of the variation in Eurozone intangible economies. This implies a relative synchronization in the cycles of Eurozone intangible intensive production activities. By comparing the decomposition of intangible economy as measured by total intangible investment and/or intangible investment not currently included in the SNA with the respective decomposition of total economy as it is measured by GVA, we can claim that intangible measurements across most Eurozone countries seems more synchronized than the aggregate business sector (e.g., compare the 1st and/or 3rd with the 5th column).

Figure 3: A global factor for Eurozone intangible economies

[5]Before applying the dynamic factor model we calculate the first differences of the log-transformed constructed variables.

Table 4: Volatility decomposition

	$\tilde{y}^{NI}_{t_q}$		$\tilde{y}^{I}_{t_q}$		$\tilde{y}^{gva}_{t_q}$	
Countries	Factor	Error	Factor	Error	Factor	Error
Austria	0.19	0.81	0.25	0.75	0.15	0.85
Belgium	0.21	0.79	0.03	0.97	0.07	0.93
Germany	0.30	0.70	0.49	0.51	0.24	0.76
Estonia	0.06	0.94	0.39	0.61	0.25	0.75
Spain	0.31	0.69	0.39	0.61	0.36	0.64
Finland	0.08	0.92	0.35	0.65	0.33	0.67
France	0.26	0.74	0.36	0.64	0.24	0.76
Ireland	0.26	0.74	0.47	0.53	0.01	0.99
Italy	0.36	0.64	0.42	0.58	0.36	0.64
Lithuania	0.33	0.67	0.03	0.97	0.12	0.88
Luxembourg	0.22	0.78	0.49	0.51	0.14	0.86
Latvia	0.03	0.97	0.22	0.78	0.18	0.82
Netherlands	0.36	0.64	0.17	0.83	0.02	0.98
Portugal	0.09	0.91	0.17	0.83	0.25	0.75
Slovenia	0.17	0.83	0.19	0.81	0.25	0.75
Slovakia	0.17	0.83	0.21	0.79	0.03	0.97

4 Conclusions

In this paper, we adopted a machine learning approach to construct quarterly time series measurements for intangible economies in Eurozone countries. Our proposed approach offers several advantages over other popular econometric techniques. It requires minimal preprocessing of the data and does not raise concerns of multicollinearity and/or endogeneity. The constructed time series can be used in various macroeconomic applications, such as macroeconomic modelling, estimation of macroeconometric models and macroeconomic policy evaluation.

Data availability

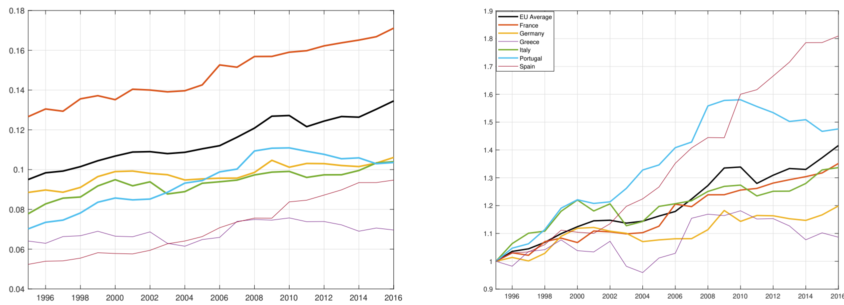
Data will become available upon request.

Acknowledgements

We would like to thank the Laboratory of Economic Policy Studies (EMOP) for the hospitality when the article was written.

Appendix

A: Additional Figures



(a) a (b) b

Figure 4: (a) GVA share of intangible investment and (b) GVA share of intangible investment (normalized 1995=1)

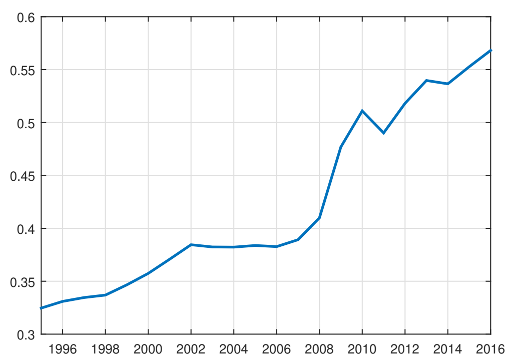


Figure 5: Intangible to tangible investment EU average

B: Technical details of data construction

By using the XGboost method (Chen and Guestrin, 2016) we wish to construct a function G such that $\hat{y}_t = G(x_t^{(1)}, \ldots, x_t^{(d)})$ is an accurate prediction for the true value y_t. XGboost is an ensemble of S regression trees (g_s, $s = 1, \ldots, S$) where a prediction $\hat{y}_t$ is obtained as

$$\hat{y}_t = G(\boldsymbol{x}_t) = \sum_{s=1}^{S} g_s(\boldsymbol{x}_t), \;\; g_s \in \mathcal{G}, \tag{2}$$

where $\boldsymbol{x}_t = (x_t^{(1)}, \ldots, x_t^{(d)})$, $\mathcal{G}$ is the space of the regression trees and each function g_s corresponds to an independent tree structure with M number of leaves and leaf scores $\boldsymbol{V} \in \mathbb{R}^M$. The functions g_s are determined by minimizing the regularized objective

$$\mathcal{L} = \sum_{t=1}^{T} \ell(\hat{y}_t, y_t) + \sum_{s=1}^{S} \Omega(g_s), \tag{3}$$

where ℓ is a differentiable convex loss function which measures the difference between the predicted and the target label and

$$\Omega(g) = M\gamma + (1/2)\lambda \sum_{m=1}^{M} v_m^2,$$

is a regularization term which penalizes the complexity of the model to avoid over-fitting with v_m being the score on the mth leaf, while γ and λ are constants that control the degree of regularization. Since the parameters of the model in (3) are the functions g_s, traditional optimization methods of the Euclidean space cannot be used. Instead the model is trained in an additive manner by first noting that the additive structure of the prediction in (2) implies that $\hat{y}_t^{(i)} = \hat{y}_t^{(i-1)} + g_i(\boldsymbol{x}_t)$, where the superscript i denotes the ith iteration of the optimization procedure. Then, objective in (3) becomes

$$\mathcal{L}^{(i)} = \sum_{t=1}^{T} \ell(y_t, \hat{y}_t^{(i-1)} + g_i(\boldsymbol{x}_t)) + \Omega(g_t). \tag{4}$$

After a second order Taylor approximation, and by removing all the constant term, it is the case that

$$\tilde{\mathcal{L}}^{(i)} = \sum_{t=1}^{T} [g_t g_i(\boldsymbol{x}_t) + \frac{1}{2} h_t g_t^2(\boldsymbol{x}_t)] + \Omega(g_i), \tag{5}$$

where $g_t = \partial_{\hat{y}^{(i-1)}} \ell(y_t, \hat{y}^{(i-1)})$ and $h_t = \partial^2_{\hat{y}^{(i-1)}} \ell(y_t, \hat{y}^{(i-1)})$ are first and second order gradient of the loss function and $\hat{y}^{(i-1)}$ is the history of the predictions up to the $(i-1)$th iteration of the algorithm. By expanding the regularization term Ω, and noting the quadratic form of equation (5), it is straightforward to find the optimal weights $\boldsymbol{V}$. Therefore, for a given tree structure we can compute the optimal leaf weights $\boldsymbol{V}$ and calculate the corresponding value of (4). Since it is impossible to make these calculations for all the possible tree structures, Chen and Guestrin (2016) show that for the loss function in (4) it is straightforward to calculate a score for a leaf node during splitting and, based on this score, they propose to utilize the so-called exact greedy algorithm in order to detect the split point that results in maximum loss reduction.

The proposed Algorithm for construction of quarterly time series data

Algorithm 1 summarizes the steps of the technique that we developed to construct quarterly time series for which we have access to their annual counterpart.

Algorithm 1 The developed algorithm for construction of quarterly time series

Input: Time series $\{y_t\}_{t=1}^T$ observed on annual frequency, D time series observed on quarterly frequency $\{x_{t_q}^{(d)}\}_{d=1}^D$, for $q = 1, 2, 3, 4$ and $t = 1, \ldots, T$.

1: Set $x_t^{(d)} = \sum_{q=1}^4 x_{t_q}^{(d)}$, for each $d = 1, \ldots, D$ and $t = 1, \ldots, T$.
2: Train XGboost on a regression model with responses $(y_1, \ldots, y_T)$ and features $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ where $\boldsymbol{x}_t = (x_t^{(1)}, \ldots, x_t^{(D)})$ to learn a function G such that $\hat{y}_t \approx y_t$ where $\hat{y}_t = G(\boldsymbol{x}_t)$ for each $t = 1, \ldots, T$.
3: Set $\hat{y}_{t_q} = G(x_{t_q}^{(1)}, \ldots, x_{t_q}^{(d)})$, for each $t = 1, \ldots, T$ and $q = 1, 2, 3, 4$.
4: **for** $t = 1, \ldots, T$ **do**
5: Set $\epsilon_t = y_t - \sum_{q=1}^4 \hat{y}_{t_q}$
6: **if** $\epsilon_t > 0$ **then**
7: Set $\tilde{y}_{t_q} = \hat{y}_{t_q} + \epsilon_t/4$
8: **else**
9: Set $\tilde{y}_{t_q} = \hat{y}_{t_q} - \epsilon_t/4$
10: **end if**
11: **end for**

Output: Time series $\{\tilde{y}_{t_q}\}_{t=1}^T$, $q = 1, 2, 3, 4$.

C: Prior distributions for the dynamic factor model

To assign prior distributions on the parameters of the dynamic factor model presented by Section 3.3 of the main paper we work as follows. To ensure identifiability of the latent factor and its loadings we follow Kose et al. (2003) and we assume that $\sigma_\eta^2 = 1$ and that $w_{11} > 0$ in equation (1) of the main paper. We also assume that σ_{ij}^2 follow inverse gamma (IG) distributions; $\sigma_{ij}^2 \overset{iid}{\sim} \text{IG}(T_Q, 0.1T_Q)$ (Berger et al., 2021), where $i = 1, \ldots, N$, $j = 1, 2, 3$ and T_Q denotes the length of each one of the constructed quarterly time series. For w_{11} we assume a truncated standard normal as prior distribution and for the rest w_{ij} parameters we set $w_{ij} \overset{iid}{\sim} N(0, 1)$. Finally, for the parameter ϕ of the autoregressive process of the latent factor we assume that it is uniformly distributed on $(-1, 1)$.

References

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area gdp. *International Journal of forecasting 20*(3), 447–460.

Ballarin, G., P. Dellaportas, L. Grigoryeva, M. Hirt, S. van Huellen, and J.-P. Ortega (2022). Reservoir computing for macroeconomic forecasting with mixed frequency data. *arXiv preprint arXiv:2211.00363*.

Berger, T., G. Everaert, and L. Pozzi (2021). Testing for international business cycles: A multilevel factor model with stochastic factor selection. *Journal of Economic Dynamics and Control 128*, 104134.

Bontadini, F., C. Corrado, J. Haskel, M. Iommi, and C. Jona-Lasinio (2023). Euklem & intanprod: industry productivity accounts with intangibles. Technical report, Luiss Lab of European Economics.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Corrado, C., J. Haskel, C. Jona-Lasinio, and M. Iommi (2016). Intangible investment in the EU and US before and since the Great Recession and its contribution to productivity growth. EIB Working Papers 2016/08, European Investment Bank (EIB).

Corrado, C., J. Haskel, C. Jona-Lasinio, and M. Iommi (2022, August). Intangible capital and modern economies. *Journal of Economic Perspectives 36*(3), 3–28.

Fiorito, R. and T. Kollintzas (1994). Stylized facts of business cycles in the g7 from a real business cycles perspective. *European Economic Review 38*(2), 235–269.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (midas): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 57–82.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. *Econometric reviews 26*(1), 53–90.

Kose, M. A., C. Otrok, and E. Prasad (2012). Global business cycles: convergence or decoupling? *International economic review 53*(2), 511–538.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country-specific factors. *American Economic Review 93*(4), 1216–1239.

Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real gdp. *Oxford Bulletin of economics and statistics 72*(1), 27–46.

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2021

01-21 Historical Cycles of the Economy of Modern Greece From 1821 to the Present, George Alogoskoufis
02-21 Greece Before and After the Euro: Macroeconomics, Politics and the Quest for Reforms, George Alogoskoufis
03-21 Commodity money and the price level, George C. Bitros. Published in: *Quarterly Journal of Austrian Economics*, 2022
04-21 Destabilizing asymmetries in central banking: With some enlightenment from money in classical Athens, George C. Bitros. Published in: *Journal of Economic Asymmetries*, 2021
05-21 Exploring the Long-Term Impact of Maximum Markup Deregulation, Athanasios Dimas and Christos Genakos
06-21 A regularization approach for estimation and variable selection in high dimensional regression models, Y. Dendramis, L. Giraitis, G. Kapetanios
07-21 Tax Competition in the Presence of Environmental Spillovers, Fabio Antoniou, Panos Hatzipanayotou, Michael S. Michael, Nikos Tsakiris
08-21 Firm Dynamics by Age and Size Classes and the Choice of Size Measure, Stelios Giannoulakis and Plutarchos Sakellaris
09-21 Measuring the Systemic Importance of Banks, Georgios Moratis, Plutarchos Sakellaris
10-21 Firms' Financing Dynamics Around Lumpy Capacity Adjustments, Christoph Görtz, Plutarchos Sakellaris, John D. Tsoukalas
11-21 On the provision of excludable public goods General taxes or user prices? George Economides and Apostolis Philippopoulos
12-21 Asymmetries of Financial Openness in an Optimal Growth Model, George Alogoskoufis
13-21 Evaluating the impact of labour market reforms in Greece during 2010-2018 , Georgios Gatopoulos, Alexandros Louka, Ioannis Polycarpou, Nikolaos Vettas
14-21 From the Athenian silver to the bitcoin standard: Private money in a state-enforced free banking model, George C. Bitros
15-21 Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities. Stelios Arvanitis and Thierry Post
16-21 Inconsistency for the Gaussian QMLE in GARCH-type models with infinite variance, Stelios Arvanitis and Alexandros Louka
17-21 Competition and Pass-Through: Evidence from Isolated Markets, Christos Genakos and Mario Pagliero
18-21 Exploring Okun's Law Asymmetry: An Endogenous Threshold LSTR Approach, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

19-21 **Limit Theory for Martingale Transforms with Heavy-Tailed Multiplicative Noise, Stelios Arvanitis and Alexandros Louka**
20-21 **Optimal taxation with positional considerations, Ourania Karakosta and Eleftherios Zacharias**
21-21 **The ECB's policy, the Recovery Fund and the importance of trust: The case of Greece, Vasiliki Dimakopoulou, George Economides and Apostolis Philippopoulos**

2022

01-22 **Is Ireland the most intangible intensive economy in Europe? A growth accounting perspective, Ilias Kostarakos, KieranMcQuinn and Petros Varthalitis**
02-22 **Common bank supervision and profitability convergence in the EU, Ioanna Avgeri, Yiannis Dendramis and Helen Louri**
03-22 **Missing Values in Panel Data Unit Root Tests, Yiannis Karavias, Elias Tzavalis and Haotian Zhang**
04-22 **Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities, Stelios Arvanitis and Thierry Post**
05-22 **Concentration Inequalities for Kernel Density Estimators under Uniform Mixing, Stelios Arvanitis**
06-22 **Public Sector Corruption and the Valuation of Systemically Important Banks, Georgios Bertsatos, Spyros Pagratis, Plutarchos Sakellaris**
07-22 **Finance or Demand: What drives the Responses of Young and Small Firms to Financial Crises? Stelios Giannoulakis and Plutarchos Sakellaris**
08-22 **Production function estimation controlling for endogenous productivity disruptions, Plutarchos Sakellaris and Dimitris Zaverdas**
09-22 **A panel bounds testing procedure, Georgios Bertsatos, Plutarchos Sakellaris, Mike G. Tsionas**
10-22 **Social policy gone bad educationally: Unintended peer effects from transferred students, Christos Genakos and Eleni Kyrkopoulou**
11-22 **Inconsistency for the Gaussian QMLE in GARCH-type models with infinite variance, Stelios Arvanitis and Alexandros Louka**
12-22 **Time to question the wisdom of active monetary policies, George C. Bitros**
13-22 **Investors' Behavior in Cryptocurrency Market, Stelios Arvanitis, Nikolas Topaloglou and Georgios Tsomidis**
14-22 **On the asking price for selling Chelsea FC, Georgios Bertsatos and Gerassimos Sapountzoglou**
15-22 **Hysteresis, Financial Frictions and Monetary Policy, Konstantinos Giakas**
16-22 **Delay in Childbearing and the Evolution of Fertility Rates, Evangelos Dioikitopoulos and Dimitrios Varvarigos**
17-22 **Human capital threshold effects in economic development: A panel data approach with endogenous threshold, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis**
18-22 **Distributional aspects of rent seeking activities in a Real Business Cycle model, Tryfonas Christou, Apostolis Philippopoulos and Vanghelis Vassilatos**

2023

01-23 **Real interest rate and monetary policy in the post Bretton Woods United States, George C. Bitros and Mara Vidali**

02-23 **Debt targets and fiscal consolidation in a two-country HANK model: the case of Euro Area, Xiaoshan Chen, Spyridon Lazarakis and Petros Varthalitis**

03-23 **Central bank digital currencies: Foundational issues and prospects looking forward, George C. Bitros and Anastasios G. Malliaris**

04-23 **The State and the Economy of Modern Greece. Key Drivers from 1821 to the Present, George Alogoskoufis**

05-23 **Sparse spanning portfolios and under-diversification with second-order stochastic dominance, Stelios Arvanitis, Olivier Scaillet, Nikolas Topaloglou**

06-23 **What makes for survival? Key characteristics of Greek incubated early-stage startup(per)s during the Crisis: a multivariate and machine learning approach, Ioannis Besis, Ioanna Sapfo Pepelasis and Spiros Paraskevas**

07-23 **The Twin Deficits, Monetary Instability and Debt Crises in the History of Modern Greece, George Alogoskoufis**

08-23 **Dealing with endogenous regressors using copulas; on the problem of near multicollinearity, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis**

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/