



Wagner's hypothesis in Europe: a causality analysis with disaggregated data

Ivan D Trofimov^{a*}

^a Kolej Yayasan Saad (KYS) Business School, Malaysia

* Corresponding author at: ivan.trofimov@kysbs.edu.my

Abstract. This paper examines Wagner hypothesis of the growth of public expenditure alongside the growth of economic activity for a panel of 28 European economies during the 1995-2018 period. The hypothesis is verified using Pesaran (2007) panel unit root and Westerlund (2007) cointegration tests that account for cross-sectional dependence in the series, and three panel causality tests (Toda-Yamamoto, Dumitrescu-Hurlin and Juodis-Karavias-Sarafidis) that are suitable for mixed order of series' integration, heterogeneous balanced panels and cases of limited evidence of cointegration. The empirical results suggested that expenditure and output variables were non-stationary in levels and stationary in the first differences; the cointegration among the variables was present; the causality was principally uni-directional (from output to public expenditure), in line with Wagner's hypothesis, or bi-directional; the causality from public expenditure to output along Keynesian lines was limited.

Keywords. Wagner's hypothesis; panel cointegration; panel causality, Europe

JEL Codes. C23; E60; H50; N44

DOI. <https://doi.org/10.17979/ejge.2023.12.1.9146>

1. Introduction

This paper empirically tests Wagner's hypothesis,¹ which assumes a faster rate of growth of government expenditure (in absolute and in relative senses) than economic growth, as well as expansion of the government spending (and, more generally, activity) at the expense of private sector (Wagner, 1912). The consideration of Wagner's hypothesis was diverse in terms of the definition and measurement of the public expenditure variable, the selection of independent variables in addition to GDP, the country coverage, econometric methods and tests, as well as the findings (the following section provides a cursory review of the empirical studies).

In contrast to many previous studies, the focus of this paper is on the relationship between output and expenditure at both aggregate and disaggregated levels. The consideration of the aggregate level of analysis is justified by the fiscal sustainability, budget deficit and public debt problems that plague

¹ In the absence of universally acceptable terms in the empirical studies, we use 'public' and 'government' expenditure terms interchangeably, and, given the conflicting findings of the empirical research, refer to Wagner's 'hypothesis', as opposed to Wagner's 'law'.



the developed economies particularly strongly (Koester & Priesmeier, 2013): the validity of Wagner's hypothesis may provide a solid theoretical explanation of these problems (in addition to other explanatory factors). On the other hand, while Wagner's hypothesis may hold at the aggregate expenditure level, this may not necessarily be the case for the individual expenditure categories, e.g. the growth of health expenditure may exceed output growth, but the growth of expenditure on agriculture may lag behind it (as put by Peacock & Scott, 2000, the references to the differing pace and speed of expansion of the government may be found in Wagner's original works). Additionally, the focus on the aggregate level may be unsound from an econometric perspective: Rajaguru (2002) and Kucukkale et al. (2012: 1-2) note the distortionary and bias-inducing effects of expenditure aggregation on cointegration and causality relationships, while Granger (1988) mentions the incongruity of the unit root properties of the aggregate and component (disaggregated) series.

Given the specific aim of the paper (to make certain generalisations as to the validity of Wagner's hypothesis in a sufficiently large group of economies) and the lack of public expenditure data with a sufficient time series dimension (that is typically available only for a limited number of economies), we use a panel data set that covers 28 European economies over the 1995-2018 period. Due to the economic and political integration that has been underway in Europe since the 1950-60s, it is reasonable to assume that the panel data would be characterized by the cross-sectional dependence (stemming principally from intra-regional trade and investment flows, implementation of common economic policies, and the public expenditure decisions made at supra-national level) and thus use appropriate panel data econometric techniques.

The rest of the paper is organized as follows. Section 2 contains a review of the theoretical basis of Wagner's hypothesis and of the relevant empirical studies. Section 3 provides the modeling framework, the data and econometric methodology, and the description of the panel data. Section 4 presents the empirical results. Section 5 concludes the paper, discusses the policy implications and outlines the avenues for future research.

2. Theoretical framework and empirical literature

Certain ambiguity exists regarding the exact definition of Wagner's hypothesis, its applicability in a specific socio-economic setting and the formulation for the modelling purposes. Peacock and Scott (2000, pp. 2-3), based on a thorough examination of Wagner's original works, note that Wagner preferred to call the concept an 'empirical observed uniformity' rather than law; tended to apply the concept both to the traditional services of the government (e.g. defence, law and order) and to the newer functions (welfare provision); included the expenditure by the central and local government as well as the activity of public enterprises; noted the problem of public sector expansion at the expense of the private sector growth; and conceptualised public sector growth not purely as a quantitative but also as a qualitative phenomenon (manifested in sophistication and greater extent of government regulation).

Timm (1961) argued that Wagner formulated the hypothesis for the analysis of the 19th-century governments but did not restrict the applicability of the concept, stating that government size would grow and involvement would extend as long as “cultural and economic progress” continues. Thus, the hypothesis may be equally applicable to the economies at the early stage of capitalist development (as were European economies in the 19th century), and to other economies at various development stages.²

The public expenditure growth at a faster rate than economic growth has been attributed to a number of objective factors. Firstly, the industrialisation, production on a large scale, and application of science and technology necessitate the accompanying growth of the organising force, embodied in government and manifested in greater spending (North & Wallis, 1982, pp. 336; Oxley, 1994, pp. 287). Secondly, in contrast to the earlier epochs (e.g. 18th and 19th centuries, when the role of the government was restricted to the provision of public order, law and basic services), the 20th century witnessed much broader roles of the government. The popularity of socialist ideas and the accompanying class struggle, the rise of central or indicative planning (along with socialist economic doctrine), the two world wars and the Cold War, the failures of the capitalist economies at certain junctures (e.g. Great Depression), led to greater public expenditure in such areas as welfare, education, health, and defense. Thirdly, the sophistication of economic activity and transactions, and scientific and technological progress required greater regulation and oversight by the government in such functional areas as the design and implementation of legislation, improvement of economic and political institutions, consumer protection from monopolies, prevention of unfair competition, and also justified greater spending to support scientific and technological capability.

The economic theory provides a number of explanations behind Wagner's hypothesis. The demand for education, health, social protection has high-income elasticity, hence the expenditure in these categories is more likely to conform to Wagner's predictions. Public spending is an outcome of a public decision making process that is subject to influence by vested interests and special interest groups through their lobbying effort, bringing the expansion of the government size and greater public spending in a number of categories (Mueller, 1987). The relative cost of government services may rise over time, resulting in spending growth in nominal terms (Baumol, 1967). Lastly, as stated by the bureaucracy theory of the government, the proliferation of the policies and the growth of spending (frequently at the expense of economic efficiency) result from the immanent features of bureaucracy and bureaucratic concerns that may be summarized as striving for power and prestige, policy and position preservation, and expansion of the bureaucratic apparatus (Niskanen, 1971; Oxley, 1994, pp. 286).

An alternative hypothesis of reverse causality that runs from government expenditure to output is rooted in Keynesian economics and is also supported by other lines of economic theory. In the short-run, during the cyclical downswing, the increase in government expenditure may stimulate output, provided that there exists idle capacity and the economy is below the potential output (Ray & Pal, 2022). This view was challenged in a number of studies: while the causality originating from government expenditure may be present, the effects on output may be negative. Barro (1991) notes the inefficiencies that are caused by the public sector expansion and that potentially slowdown economic

² An alternative argument is the exclusive applicability of Wagner's hypothesis to the society that undergo rapid industrialisation, and not to the pre- or post-industrial social settings (Jaen Garcia, 2004, p. 13).

growth. Tobin (2005) argues that the strength and sign of the expenditure-output effect hinges on the specific role of the state as a modernisation agent (as opposed to a product of vested interests and political pressures, or a source of bureaucratic failure and slack). The positive effects of expenditure on output are, therefore, not guaranteed.

The empirical testing of Wagner's hypothesis was diverse in many respects. With regard to the definition of government expenditure, the empirical studies chose between the nominal and real expenditure values (for instance, Gandhi, 1971, and Lin, 1995, are examples of the study that used expenditure in current prices, while Bohl, 1996, is the study that defined expenditure in constant prices). In a number of studies, the defence expenditure was excluded to isolate the expenditure-output relationships in a civil economy (Abizadeh, Gray, 1985). The empirical research focused on both the expenditure by the central government (Mann, 1980) and the spending at sub-national level (e.g. Narayan et al, 2012 in the study of Wagner's hypothesis for a group of Indian states), but tended, in the absence of disaggregated expenditure data, to examine the aggregate expenditure (few studies, such as the analysis of the hypothesis in Canada by Biswal et al, 1999, stand as exception).

Regarding the selection of regressors, the following variables were used in addition to GDP (per capita): the sectoral shares of output to account for structural influences on 'expenditure-output' nexus (Mann, 1980), the values of exports and imports to account for external economy influences of government spending decisions (Abizadeh, Gray, 1985); permanent income and deviation of current demand from the trend (Courakis et al, 1993); country size and population density (Alesina, Wacziarg, 1998; Dao, 1995).

In terms of geographic coverage, a large number of studies focused on a single economy: Wagner's hypothesis in USA (Islam, 2001), Spain (Jaen Garcia, 2004), Italy (Magazzino, 2012), Canada (Ahsan et al, 1996), Egypt (Ghazy et al, 2021), among others. Selected studies looked at regional or level of development groups (developing economies by Diamond, 1977; 14 European economies by Afonso, Alves, 2017), economies with specific characteristics (e.g. the study of the hypothesis for petroleum exporters, where oil export revenues constitute a major source of government revenue and thus influence fiscal policy significantly, Burney, 2002), or random sets of economies (e.g. the comparative study of Spain and Armenia by Sedrakyan and Varela-Candamio, 2019; or of the three African economies, Ansari et al, 1997).

In terms of econometric methods and tests, the earlier studies tended to rely on descriptive statistics analysis (Bairam, 1992), the use of linear regression (Lin, 1995), or specification of the simultaneous equation models for the demand and supply of public expenditure (Dollery, Singh, 2000). Such methods suffered from the 'spurious regression' and identification problems and ignored the non-stationarity of the time series. To address these problems, a more recent empirical analysis started to utilise cointegration and causality tests (Oxley, 1994; Legrenzi, 2000), vector autoregression models (Benavides et al, 2013), panel data techniques (Afonso, Jalles, 2014), or non-linear models (Karagianni, Pempetzoglou, 2009).

The empirical research delivered conflicting results. A number of studies indicated the absence of causality between expenditure and GDP in either direction (Huang, 2006; Sinha, 2007). The unidirectional causality that supports Wagner's hypothesis was identified in the studies by Courakis et

al (1993), Tang (2001), Chang et al (2004), Jaen Garcia (2004), Sideris (2007), Lamartina and Zaghini (2011) and, in selected specifications, Benavides et al (2013). The causality along Keynesian lines was demonstrated by Iyare and Lorde (2004) and Babatunde (2011). The bi-directional causality was indicated by Narayan et al (2008), Ziramba (2009), and Yay and Tastan (2009). Given a variety of empirical outcomes in the studies that focused at aggregate expenditure and individual economies (or narrow groups of economies), it is advantageous to consider the specific expenditure categories and a broader panel that encompassed the maximum possible number of countries.

3. Methodology

Model

Following Loizides and Vamvoukas (2005), a bivariate representation of Wagner's hypothesis is used, with no inclusion of additional variables. The five specifications of the hypothesis suggested by Peacock and Wiseman (1979), Mann (1980), Musgrave (1969), Gupta (1967), and Goffman (1968) were considered:

$$\text{Specification 1 } \ln G = \alpha + \beta \ln Y + \varepsilon_t \quad (1)$$

$$\text{Specification 2 } \ln(G/Y) = \alpha + \chi \ln Y + \varepsilon_t \quad (2)$$

$$\text{Specification 3 } \ln(G/Y) = \alpha + \delta \ln(Y/P) + \varepsilon_t \quad (3)$$

$$\text{Specification 4 } \ln(G/P) = \alpha + \phi \ln(Y/P) + \varepsilon_t \quad (4)$$

$$\text{Specification 5 } \ln G = \alpha + \varphi \ln(Y/P) + \varepsilon_t \quad (5)$$

where G , Y and P respectively represent government, expenditure, real GDP and population.

It is assumed that causality runs from the right to the left-hand side of the above equations under Wagner's hypothesis, and in the opposite direction under Keynesian hypothesis.

Data

The paper uses the GDP and public expenditure data released by Eurostat. Both GDP and expenditure were represented in millions of Euro in chain-linked volume measures of GDP (with the base of the chain-linked volume index set at 2010). The constructed panel dataset used in the empirical analysis covered 1995-2018 period and included 28 European economies: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain,

Sweden, Switzerland, and the UK.

The public expenditure categories followed Classification of the Function of the Government (COFOG), Version 1999 (OECD, 2011, pp. 194-5). In addition to total expenditure, the following expenditure categories are delineated: defence; economic affairs; education; environmental protection; general public services; health; housing and community amenities; public order and safety; recreation, culture and religion; social protection.

Econometric methods

Initially, we examine the integration order and the unit root properties of the expenditure and output. There is a possibility of cross-sectional dependence in the series (the correlation of the panel members in the cross-section due to various unobserved common factors, such as policy and political integration or policy learning). As a result, we conduct Pesaran (2004, 2015) cross-sectional dependence test that is flexible with regard to panel data structure (specifically when time series dimension is smaller than cross-sectional dimension, $T < N$), and Pesaran (2007) cross-sectionally augmented IPS test (CIPS) that is robust to the cross-sectional correlation (in lieu of the first-generation panel unit root tests that ignore the phenomenon).

For the variables that have $I(1)$ integration order and contain unit roots, we consider the possibility of cointegration and conduct Westerlund panel cointegration tests (Westerlund, 2007). Westerlund test tackles cross-sectional dependence, removes common factor restriction, and specifies the error-correction model as follows (Persyn & Westerlund, 2008, pp. 233):

$$\Delta y_{it} = \delta'_i d_t + \alpha_i (y_{it-1} - \beta' z_{it-1}) + \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{it-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta z_{it-j} + \varepsilon_{it}, \quad (6)$$

where d_t is a deterministic component, and α_i is a coefficient that measures the speed at which the system returns to the equilibrium ($y_{it-1} - \beta' z_{it-1}$). The null hypothesis is of the absence of error-correction and cointegration ($\alpha_i = 0$ for all i); an alternative hypothesis is of the presence of error correction and cointegration for some of i , or all i 's ($\alpha_i < 0$ in the test's group-mean statistics or $\alpha_i = \alpha < 0$ in the panel statistics).

As a next step, in order to ascertain causality between public expenditure and output, we employed Toda-Yamamoto (TY) causality test (Toda & Yamamoto, 1995). The test is applicable for the situations when the variables have different order of integration, for instance, the combination of $I(0)$ and $I(1)$ order (the data characteristic that renders the conventional Granger causality test unreliable), or when the cointegration tests give conflicting indications. The TY test is invariant to the integration orders (allows any combination of the orders) and is valid irrespective of the presence or absence of cointegration. Implementation-wise, the TY test establishes the maximum integration order of the series, d_{\max} , uses it in the estimation of the augmented VAR model in levels (where the optimal lag

order is determined using the conventional lag selection criteria, and the selected lag for each variable is increased by d_{\max}) and constructs modified Wald (*MWald*) statistics to confirm causality (similarly to the conventional Granger causality test). The inclusion of lags is necessary, due to the delayed (not instantaneous) effects of GDP on government expenditure.

For the case of public expenditure-output relationship, the augmented VAR model is represented as:

$$Y_t = \alpha_0 + \sum_{i=1}^k \phi_{1i} Y_{t-i} + \sum_{j=k+1}^{d_{\max}} \phi_{1j} Y_{t-j} + \sum_{j=1}^k \gamma_{1i} X_{t-j} + \sum_{j=k+1}^{d_{\max}} \lambda_{1i} X_{t-j} + \varepsilon_{1t} \quad (7)$$

$$X_t = \alpha_1 + \sum_{i=1}^k \phi_{2i} X_{t-i} + \sum_{j=k+1}^{d_{\max}} \phi_{2j} X_{t-j} + \sum_{j=1}^k \gamma_{2i} Y_{t-j} + \sum_{j=k+1}^{d_{\max}} \lambda_{2i} Y_{t-j} + \varepsilon_{2t} \quad (8)$$

The causality in the sense of Granger from X to Y in Equation (7) exists when $\gamma_{1i} \neq 0, \forall i$, and the like causality from Y to X in Equation (8) is present when $\gamma_{2i} \neq 0, \forall i$.

For the purpose of robustness check, we conducted Dumitrescu-Hurlin (2012) test of causality in a panel of stationary data (i.e. to implement the test the $I(1)$ variable were converted to the first difference form). The estimated coefficients are time-invariant but are allowed to differ across the panel members (the test is thus suitable for heterogeneous balanced panels, Lopez & Weber, 2017, pp. 3-4). The lag order (K) that is uniform across panel members is determined using the usual information criteria, with the first K_{\max} eliminated during the lag selection process.

For the case of two stationary variables, the Dumitrescu-Hurlin model is set as (Lopez & Weber, 2017, pp. 3-4):

$$X_{it} = \alpha_i + \sum_{k=1}^K \delta_{ik} X_{it-k} + \sum_{k=1}^K \gamma_{ik} Y_{it-k} + \varepsilon_{it}, \quad (9)$$

for $i=1, \dots, N$ and $t=1, \dots, T$.

The null hypothesis is of no causality for any cross-sections in the panel:

$$H_0 : \gamma_{i1} = \dots = \gamma_{iK} = 0 \text{ for } \forall i = 1, \dots, N \quad (10)$$

Under an alternative hypothesis the causality in some of the cross-sections is allowed:

$$H_A : \gamma_{i1} = \dots = \gamma_{iK} = 0 \text{ for } \forall i = 1, \dots, N_1 \text{ and} \quad (11)$$

$$\gamma_{i1} \neq 0 \text{ or } \dots \gamma_{iK} \neq 0 \text{ for } \forall i = N_1 + 1, \dots, N \quad (12)$$

For each i , the Wald test is conducted for the hypothesis $\gamma_{iK} = 0$ and the Wald test statistics for each panel member and the aggregate panel are obtained as:

$$W_{it} \text{ and } \bar{W} = 1 / N \sum_{i=1}^N W_{it} \tag{13}$$

As a last step the standardized Z and \tilde{Z} statistics are calculated:

$$Z = \sqrt{\frac{N}{2K}} (\bar{W} - K) \xrightarrow[T, N \rightarrow \infty]{d} N(0,1) \text{ and} \tag{14}$$

$$\tilde{Z} = \sqrt{\frac{N}{2K} \cdot \frac{T-3K-5}{T-2K-3}} \cdot \left[\frac{T-3K-3}{T-3K-1} \cdot \bar{W} - K \right] \xrightarrow[N \rightarrow \infty]{d} N(0,1) \tag{15}$$

As a last step we perform the causality test proposed by Juodis et al. (2021). In contrast to Dumitrescu-Hurlin test that has the highest power with large T dimension and N/T^2 approaching zero, it is suitable for the panels with a moderate time series dimension. The test performs well in both homogeneous and heterogeneous panels, assumes that Granger causality parameters are equal to zero under the null, and thus allows using pooled estimator for Granger causality parameters (Xiao et al., 2021, p. 4). The pooled estimators have faster rate of convergence while at the same time suffering from the ‘Nickel bias’, the problem corrected in the test using the half-panel jackknife method (Xiao et al., 2021, p. 3).

The test equation is given as (Xiao et al., 2021, p. 3):

$$X_{it} = \phi_{0i} + \sum_{p=1}^P \phi_{pi} X_{it-p} + \sum_{p=1}^P \gamma_{pi} Y_{it-p} + \varepsilon_{it}, \tag{16}$$

where $i = 1, \dots, N$, $t = 1, \dots, T$, $p = 1, \dots, P$, ϕ_{0i} are individual specific effects, ϕ_{pi} are heterogeneous autoregressive coefficients, and γ_{pi} are heterogeneous feedback coefficients. Y_{it} is assumed to be a scalar. The null hypothesis assumes the absence of Granger causality $H_0 : \gamma_{pi} = 0$ for all i and p , while the alternative hypothesis assumes that $\gamma_{pi} \neq 0$ for some i and p .

4. Empirical results

Figure 1 (Appendix) shows the dynamics of cross-sectional means of GDP and expenditure series. Both GDP and GDP per capita exhibited an upward trend throughout the period (with the exception of the global financial crisis / GFC years of 2008-09). With regard to absolute values of expenditures, the upward trend during the whole period (1995-2018) was observed for the total, education, health, public order and safety, social protection and recreation, culture and religion categories, while other categories were characterised by stabilisation of expenditure (economic affairs, general public services, and environmental protection), decline (defence and housing and community amenities) in the post-GFC years. The expenditure shares of GDP did not follow any specific trend (possibly with the exception of defence, and general public services). For the expenditure per capita, the deterministic patterns were observed only for education, health, public order and safety, social protection and recreation, culture and religion categories. The visual inspection is supplemented for formal unit root tests; the former method is deemed misleading (Cuddington et al, 2002: 21).

As demonstrated in Table 1, according to the Pesaran 2004 and 2015 tests, the null hypotheses of cross-sectional independence and weak dependence were respectively rejected: each of the variables in question was found to be characterised by strong cross-sectional correlation. The pattern is not unusual, given the high degree of economic and political integration in Europe: the previous research indicated convergence in public expenditure across European economies (Apergis et al., 2013) and also highlighted GDP and business cycle comovement in the region (Azcona, 2022). The incorporation of the cross-sectional dependence into unit root and cointegration tests is therefore warranted.

The results of the Pesaran CADF unit root test are presented in Table 1. The test was implemented with a trend component in order to account for the growth of public expenditure and GDP in the long-run. The test was also run on the levels and the first differences of the variables in order to determine the integration order of the variables, which is instrumental for Toda-Yamamoto causality test (that requires the knowledge of the maximum order of integration). The test indicates that GDP and GDP per capita variables were trend stationary in both levels and the first differences (in the latter case at 5% and 10% significance levels), and are thus $I(0)$ order variables. All other variables in the level form contained unit root and the test' null hypothesis was not rejected (the exceptions were the absolute level of total, education, health, public order and safety, environmental protection and housing and community amenities expenditure; the per capita expenditure on education, public order and safety; and the public expenditure as a proportion of GDP in general public services and environmental protection). In the first difference form, all expenditure variables were trend stationary at 5% level (social protection expenditure as a proportion of GDP at 10% level). Overall, we conclude that the variables are the mix of $I(0)$ and $I(1)$ order of integration, and the maximum order is $I(1)$.

The Westerlund cointegration test was implemented with a maximum of two lags, one lead and a short kernel window of two (in line with recommendations for panels with small time-series dimension, Persyn & Westerlund, 2008) for each of the five specifications of Wagner's hypothesis. Alternative lag-lead and kernel width combinations were also tried, yielding rather similar results (results available on request). A total of 220 test statistics were obtained: four statistics, (G_t, G_a, P_t and P_a) for five specifications (Peacock-Wiseman, Mann, Musgrave, Gupta and Goffman), and 11 expenditure categories. The test failed to reject the null hypothesis of no cointegration in 41 instances (i.e. by 41 statistics), thus suggesting that overall the public expenditure and GDP were cointegrated. The greatest number of non-rejections was indicated in the social protection expenditure category, suggesting that a long-term equilibrium relationship between public expenditure and GDP was less likely in this category. Out of four Westerlund test statistics, the greatest number of non-rejections of the null was indicated for the G_a statistics. Out of five specifications, Mann and Musgrave models (Specifications 2 and 3) gave somewhat weaker support for cointegration.

While the evidence of cointegration relationships is generally strong, we relied on Toda-Yamamoto causality test that is invariant to the presence of the long-run equilibrium relationship and the (non-) stationarity of the variables. The Pesaran CADF unit root test unequivocally indicated the highest order of integration of one. The optimal lag length for the test was selected based on a combination of information criteria (Akaike, Schwarz-Bayesian and Hannan-Quinn). Given that the too-short lag length

causes bias, while longer than optimal lag length results in test' inefficiency (Caporale & Pittis, 1999; Ayad & Belmokaddem, 2017) we considered alternative lag structures and, where applicable, reported the results with the highest and the second highest lag length. The Toda-Yamamoto test was run for 100 cases (given five possible specifications of Wagner's hypothesis and alternative lag structures for the test). The causality running unidirectionally from public expenditure to GDP was not indicated in a single case, while bi-directional causality was noted in 37 cases (Table 3). No causality in either direction was demonstrated in 14 cases, while causality from GDP to public expenditure in line with Wagner's hypothesis was shown in 49 cases. Overall, Toda-Yamamoto test does not support the Keynesian hypothesis of the uni-directional influence of public expenditure on GDP.

As part of the robustness check, we conducted Dumitrescu-Hurlin causality test on each specification, with variables expressed as first-differences. The optimal lag length was determined by Bayesian-Schwarz criterion from a maximum of eight lags, and in all cases was found to be one. Table 4 contains an asymptotically well-behaved average Wald statistics (\bar{W}), the standardised Z and approximated standardised \tilde{Z} statistics that follow normal distribution (the latter used for the panels with fixed time-series dimension), together with the respective p-values. We note that in a number of instances Z and \tilde{Z} statistics gave conflicting results; however, based on \tilde{Z} , the test rejected the null hypothesis that public expenditure does not Granger-cause GDP only in five cases (three specifications for the general public services category, and two specifications for the defence category). In contrast, the null hypothesis that GDP does not Granger-cause public expenditure was rejected in a total of 26 cases, while bi-directional causality was indicated in 7 cases.

Lastly, given the relatively short N and T dimensions of the panel, we implemented the bias-corrected Granger causality test proposed by Xiao et al. (2021). The test was run on the first differences of the series with optimal lag selected based on Bayesian-Schwarz criterion (similarly to the previous test, in all cases, the optimal lag length stood at one). The results are presented in Table 5: the half-panel Jackknife (HPJ) Wald test statistics and the estimator, both with the respective p-values. The test rejected the null of no unidirectional causality from GDP to public expenditure in a total of 29 out of 55 cases, but rejected the null of no unidirectional causality from public expenditure to GDP in only 2 cases (Mann and Musgrave specifications for the public order and safety category). The bi-directional causality was indicated in 11 cases.

Robustness checks

In addition to causality testing, the study has experimented with cross-sectional analysis of the data. For each year of the period (1995-2018), the cross-sectional average was calculated for each output and expenditure variable. The Wagner's hypothesis was then verified through the significance of the coefficient in the regression that includes expenditure as independent variable and GDP (output) as regressor. Table 6 in the Appendix demonstrates the coefficient significance in most expenditure categories in Specifications 1 and 4, in the majority of categories in Specification 5, and in many categories in other specifications.

The Toda-Yamamoto causality results were consistent over time. Two checks that capture the effect of the business cycle on the causality and that distinguish causality in the expansion and downturn were implemented. The first check introduced the cyclical dummy variable in the Toda-Yamamoto relationship (the dummy takes the value one when GDP has negative growth and zero value when the growth rate is positive). As indicated in Table 7 in the Appendix, the inclusion of the dummy variable did not alter the results substantially, with bi-directional causality replacing Wagner hypothesis causality in few categories (specifications) and vice versa. Importantly, the Keynesian-type causality from expenditure to output was not identified in any of the cases, similar to the baseline model. The second check included the specification of output in expansionary and recessionary stages (the calculation of the growth rate of GDP, the separation of instances of positive and negative growth, and calculation of the cumulative sums of positive and negative growth). The check was performed for Specification 1, for the whole period and the two sub-periods prior to and after GFC (the maximum lags were tried only for the whole period, given a limited number of observations in the sub-periods). For the whole period, bi-directional causality was identified in 10 out of 40 cases, while Wagner type causality was observed in 25 cases (Table 8 in the Appendix). For the post-GFC period, the number of Wagner type causality cases stood at 11 out of 22, and for the pre-GFC period at 9 out of 22. Bi-directional or no causality were observed in other instances. Keynesian causality was identified only in two instances in pre-GFC period, however at higher lags this causality does not hold. Overall, the robustness checks give results that are very similar to the baseline model.

Table 1. Pesaran CD test and CADF test results.

Expenditure area	Variable	Pesaran (2004) CD test		Pesaran (2015) CD test		CADF (trend, levels)		CADF (trend, first differences)	
		Stat	p-value	Stat	p-value	Stat	p-value	Stat	p-value
GDP	Y	95.240	0.000	85.585	0.000	-3.124	0.000	-2.629	0.035
GDP per capita	Y/P	95.232	0.000	81.140	0.000	-3.080	0.000	-2.583	0.060
Total	G/Y	95.214	0.000	24.075	0.000	-2.693	0.015	-3.424	0.000
Total	G/P	95.229	0.000	52.590	0.000	-2.789	0.003	-3.642	0.000
Total	G	95.236	0.000	83.761	0.000	-2.847	0.001	-3.880	0.000
Economic affairs	G/Y	93.318	0.000	8.802	0.000	-2.523	0.113	-4.011	0.000
Economic affairs	G/P	95.131	0.000	14.079	0.000	-2.512	0.126	-3.906	0.000
Economic affairs	G	95.175	0.000	34.250	0.000	-2.698	0.014	-4.003	0.000
Defense	G/Y	33.609	0.000	27.440	0.000	-2.286	0.555	-3.846	0.000
Defense	G/P	95.057	0.000	8.194	0.000	-2.379	0.348	-4.086	0.000
Defense	G	95.124	0.000	5.827	0.000	-2.598	0.051	-4.141	0.000
Education	G/Y	95.045	0.000	16.610	0.000	-2.262	0.608	-3.588	0.000
Education	G/P	95.208	0.000	48.748	0.000	-2.844	0.001	-3.461	0.000
Education	G	95.228	0.000	71.561	0.000	-2.997	0.000	-3.652	0.000
Health	G/Y	94.694	0.000	46.563	0.000	-2.572	0.068	-3.110	0.000
Health	G/P	95.188	0.000	76.444	0.000	-2.634	0.033	-3.166	0.000
Health	G	95.214	0.000	82.476	0.000	-2.940	0.000	-3.431	0.000
Public order and safety	G/Y	64.675	0.000	4.521	0.000	-2.331	0.452	-3.263	0.000
Public order and safety	G/P	95.177	0.000	51.373	0.000	-2.825	0.002	-3.436	0.000
Public order and safety	G	95.203	0.000	70.858	0.000	-3.029	0.000	-3.762	0.000
Social protection	G/Y	95.121	0.000	27.647	0.000	-2.064	0.919	-2.584	0.060
Social protection	G/P	95.220	0.000	70.212	0.000	-2.183	0.765	-2.911	0.000
Social protection	G	95.230	0.000	88.382	0.000	-1.948	0.980	-2.752	0.006
General public services	G/Y	94.471	0.000	29.048	0.000	-2.989	0.000	-3.819	0.000
General public services	G/P	95.166	0.000	1.752	0.080	-2.545	0.090	-3.478	0.000
General public services	G	95.201	0.000	17.338	0.000	-2.606	0.046	-3.539	0.000
Environmental protection	G/Y	42.025	0.000	5.969	0.000	-2.815	0.002	-3.851	0.000
Environmental protection	G/P	94.405	0.000	21.722	0.000	-2.762	0.005	-3.563	0.000
Environmental protection	G	94.859	0.000	40.309	0.000	-2.840	0.001	-3.910	0.000
Recreation, culture, religion	G/Y	9.367	0.000	8.040	0.000	-2.278	0.572	-3.388	0.000
Recreation, culture, religion	G/P	95.066	0.000	36.176	0.000	-2.556	0.080	-3.372	0.000
Recreation, culture, religion	G	95.164	0.000	57.551	0.000	-2.707	0.012	-3.421	0.000
Housing and community amenities	G/Y	29.495	0.000	18.457	0.000	-2.757	0.005	-3.740	0.000
Housing and community amenities	G/P	94.292	0.000	5.234	0.000	-2.735	0.008	-3.860	0.000
Housing and community amenities	G	94.905	0.000	6.287	0.000	-2.868	0.001	-3.736	0.000

Note. All variables are in the logarithm form. The Pesaran CADF test critical values at 10%, 5% and 1% levels are -2.580, -2.660, and -2.810 respectively.

Table 2. Westerlund cointegration test results.

Expenditure area	Variable	Specification 1 Peacock-Wiseman		Specification 2 Mann		Specification 3 Musgrave		Specification 4 Gupta		Specification 5 Goffman	
		D(G) D(Y)		D(G/Y) D(Y)		D(G/Y) D(Y/P)		D(G/P) D(Y/P)		D(G) D(Y/P)	
Total	Gt	-3.147	0.000	-3.147	0.000	-3.108	0.000	-2.593	0.000	-2.874	0.000
	Ga	-12.747	0.000	-6.083	0.848	-6.272	0.801	-8.719	0.063	-11.625	0.000
	Pt	-15.051	0.000	-14.431	0.000	-14.502	0.000	-14.359	0.000	-13.517	0.000
	Pa	-10.429	0.000	-5.479	0.068	-5.531	0.061	-9.286	0.000	-8.156	0.000
Education	Gt	-2.385	0.000	-2.385	0.000	-2.532	0.000	-2.251	0.003	-2.235	0.004
	Ga	-10.305	0.001	-5.272	0.966	-5.993	0.868	-10.102	0.002	-8.833	0.050
	Pt	-10.370	0.003	-10.191	0.006	-11.235	0.000	-10.849	0.001	-11.435	0.000
	Pa	-8.331	0.000	-5.207	0.122	-5.845	0.027	-8.976	0.000	-7.918	0.000
Economic affairs	Gt	-2.636	0.000	-2.636	0.000	-2.640	0.000	-2.676	0.000	-2.646	0.000
	Ga	-11.200	0.000	-9.484	0.011	-10.038	0.002	-11.439	0.000	-10.633	0.000
	Pt	-15.750	0.000	-15.670	0.000	-15.946	0.000	-15.404	0.000	-15.369	0.000
	Pa	-11.689	0.000	-11.182	0.000	-11.605	0.000	-12.243	0.000	-10.838	0.000
Defense	Gt	-2.324	0.000	-2.324	0.001	-2.228	0.004	-2.153	0.014	-2.185	0.008
	Ga	-6.541	0.721	-7.666	0.305	-7.063	0.531	-6.815	0.625	-6.973	0.566
	Pt	-11.095	0.000	-11.142	0.000	-11.014	0.000	-10.213	0.005	-10.457	0.003
	Pa	-7.071	0.000	-6.982	0.001	-6.848	0.001	-6.220	0.009	-6.608	0.002
Health	Gt	-3.021	0.000	-3.021	0.000	-3.041	0.000	-2.597	0.000	-2.744	0.000
	Ga	-15.795	0.000	-7.471	0.375	-6.924	0.584	-11.931	0.000	-12.402	0.000
	Pt	-14.370	0.000	-12.472	0.000	-12.773	0.000	-13.233	0.000	-13.933	0.000
	Pa	-13.334	0.000	-9.349	0.000	-8.813	0.000	-10.820	0.000	-11.121	0.000
Social protection	Gt	-2.630	0.000	-2.630	0.000	-2.522	0.000	-2.408	0.000	-2.489	0.000
	Ga	-6.039	0.858	-2.905	1.000	-2.841	1.000	-6.099	0.845	-5.663	0.925
	Pt	-10.889	0.001	-9.029	0.084	-8.594	0.174	-11.481	0.000	-10.817	0.001
	Pa	-6.109	0.013	-3.283	0.871	-2.876	0.947	-5.033	0.169	-4.681	0.296

Table 2. (Cont).

Public order and safety	Gt	-2.593	0.000	-2.593	0.000	-2.566	0.000	-2.089	0.034	-2.618	0.000
	Ga	-9.733	0.006	-7.253	0.457	-7.390	0.405	-7.565	0.341	-9.064	0.031
	Pt	-12.577	0.000	-13.962	0.000	-13.940	0.000	-11.949	0.000	-12.971	0.000
	Pa	-7.873	0.000	-7.909	0.000	-7.641	0.000	-6.629	0.002	-7.326	0.000
General public services	Gt	-2.369	0.000	-2.369	0.000	-2.192	0.007	-2.113	0.024	-2.178	0.009
	Ga	-7.489	0.368	-6.942	0.577	-5.688	0.921	-7.175	0.487	-6.468	0.744
	Pt	-12.926	0.000	-12.820	0.000	-12.971	0.000	-12.342	0.000	-12.845	0.000
	Pa	-8.278	0.000	-7.714	0.000	-7.800	0.000	-8.006	0.000	-8.067	0.000
Environmental protection	Gt	-2.339	0.001	-2.339	0.001	-2.464	0.000	-2.357	0.000	-2.608	0.000
	Ga	-8.991	0.036	-8.058	0.187	-8.336	0.123	-8.523	0.090	-9.969	0.003
	Pt	-11.330	0.000	-11.317	0.000	-10.769	0.001	-12.300	0.000	-11.051	0.000
	Pa	-7.516	0.000	-6.883	0.001	-6.393	0.005	-8.480	0.000	-7.053	0.000
Recreation, culture, religion	Gt	-2.001	0.094	-2.001	0.094	-2.103	0.028	-1.862	0.310	-2.278	0.002
	Ga	-8.151	0.164	-5.426	0.952	-5.975	0.872	-7.149	0.497	-8.488	0.096
	Pt	-10.135	0.006	-10.099	0.007	-10.134	0.006	-9.094	0.075	-10.454	0.003
	Pa	-7.613	0.000	-6.229	0.009	-6.541	0.003	-5.995	0.018	-7.857	0.000
Housing and community amenities	Gt	-2.862	0.000	-2.862	0.000	-2.615	0.000	-2.588	0.000	-2.848	0.000
	Ga	-8.446	0.103	-8.193	0.154	-8.154	0.163	-7.591	0.331	-8.551	0.085
	Pt	-12.283	0.000	-12.329	0.000	-11.711	0.000	-13.035	0.000	-12.140	0.000
	Pa	-8.131	0.000	-7.855	0.000	-7.664	0.000	-9.089	0.000	-8.204	0.000

Note. The statistics in bold indicates non-rejection of the null hypothesis.

Table 3. Toda-Yamamoto panel causality test results

Expenditure area		Specification 1 Peacock-Wiseman		Specification 2 Mann		Specification 3 Musgrave		Specification 4 Gupta		Specification 5 Goffman	
		D(G) D(Y)	D(Y)	D(G/Y) D(Y)	D(Y) D(G/Y)	D(G/Y) D(Y/P)	D(Y/P) D(G/Y)	D(G/P) D(Y/P)	D(Y/P) D(G/P)	D(G) D(Y/P)	D(Y/P) D(G)
1	Chi square	123.284	27.356	45.688	27.356	40.079	24.389	33.964	0.281	86.866	28.584
	Prob	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.869	0.000	0.000
	Lag	4	4	4	4	4	4	2	2	4	4
		Bi-directional		Bi-directional		Bi-directional		Bi-directional		Bi-directional	
	Chi square	133.328	23.566	54.452	23.566	48.766	23.720	68.708	13.088	98.053	24.664
	Prob	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000
2	Lag	5	5	5	5	5	5	4	4	5	5
		Bi-directional		Bi-directional		Bi-directional		Bi-directional		Bi-directional	
	Chi square	40.058	10.212	31.179	10.212	31.835	10.133	31.660	7.525	29.158	16.343
	Prob	0.000	0.037	0.000	0.037	0.000	0.038	0.000	0.111	0.000	0.003
	Lag	4	4	4	4	4	4	4	4	4	4
		Bi-directional		Bi-directional		Bi-directional		Y→G		Bi-directional	
3	Chi square	15.608	1.191	5.734	1.191	5.491	1.451	12.011	3.187	13.582	1.351
	Prob	0.001	0.755	0.125	0.755	0.139	0.694	0.003	0.203	0.004	0.717
	Lag	3	3	3	3	3	3	2	2	3	3
		Y→G		No causality		No causality		Y→G		Y→G	
	Chi square	39.619	9.280	35.848	9.280	34.014	11.722	10.529	2.748	38.216	10.650
	Prob	0.000	0.233	0.000	0.233	0.000	0.110	0.015	0.432	0.000	0.222
4	Lag	7	7	7	7	7	7	3	3	8	8
		Y→G		Y→G		Y→G		Y→G		Y→G	
	Chi square	106.126	0.847	18.663	0.847	18.242	1.757	83.357	5.503	76.628	1.211
	Prob	0.000	0.932	0.001	0.932	0.001	0.780	0.000	0.139	0.000	0.876
	Lag	4	4	4	4	4	4	3	3	4	4
		Y→G		Y→G		Y→G		Y→G		Y→G	
4	Chi square							65.742	1.974		
	Prob							0.000	0.741		
	Lag							4	4		
							Y→G				

Table 3. (Cont).

5	Chi square	116.408	11.018	38.008	11.018	32.330	10.884	61.118	10.971	97.801	11.314	
	Prob	0.000	0.026	0.000	0.026	0.000	0.028	0.000	0.027	0.000	0.023	
	Lag	4	4	4	4	4	4	4	4	4	4	
		Bi-directional		Bi-directional		Bi-directional		Bi-directional		Bi-directional		
	Chi square	99.358	21.883	50.920	21.883	38.204	12.770	63.230	15.889	71.674	12.500	
	Prob	0.000	0.005	0.000	0.005	0.000	0.047	0.000	0.044	0.000	0.052	
	Lag	8	8	8	8	6	6	8	8	6	6	
		Bi-directional		Bi-directional		Bi-directional		Bi-directional		Bi-directional		
	6	Chi square	24.435	0.384	0.575	0.384	7.317	0.735	18.662	0.013	9.018	0.258
		Prob	0.000	0.825	0.750	0.825	0.026	0.692	0.000	0.994	0.011	0.879
Lag		2	2	2	2	2	2	2	2	2	2	
		Y→G		No causality		Y→G		Y→G		Y→G		
Chi square		50.145	4.343	10.492	4.343	5.750	7.161	34.687	4.209	33.511	4.038	
Prob		0.000	0.362	0.033	0.362	0.219	0.128	0.000	0.520	0.000	0.401	
Lag		4	4	4	4	4	4	5	5	4	4	
		Y→G		Y→G		No causality		Y→G		Y→G		
7		Chi square	107.205	10.899	13.445	10.899	26.908	10.355	75.663	5.104	123.616	16.208
		Prob	0.000	0.012	0.004	0.012	0.000	0.066	0.000	0.403	0.000	0.006
	Lag	3	3	3	3	5	5	5	5	5	5	
		Bi-directional		Bi-directional		Bi-directional		Y→G		Bi-directional		
	Chi square	155.005	12.198	32.839	12.198	9.300	10.229	50.810	1.501	88.522	14.007	
	Prob	0.000	0.032	0.000	0.032	0.026	0.017	0.000	0.682	0.000	0.003	
	Lag	5	5	5	5	3	3	3	3	3	3	
		Bi-directional		Bi-directional		Bi-directional		Y→G		Bi-directional		
	8	Chi square	1.720	1.124	23.417	1.124	25.006	0.831	2.024	0.612	1.484	0.838
		Prob	0.423	0.570	0.000	0.570	0.000	0.660	0.363	0.737	0.476	0.658
Lag		2	2	2	2	2	2	2	2	2	2	
		No causality		Y→G		Y→G		No causality		No causality		
Chi square		2.611	5.814	13.948	5.814	15.238	4.550	1.289	2.225	2.615	4.468	
Prob		0.456	0.121	0.003	0.121	0.002	0.208	0.732	0.527	0.455	0.215	
Lag		3	3	3	3	3	3	3	3	3	3	
		No causality		Y→G		Y→G		No causality		No causality		

Table 3. (Cont).

9	Chi square	17.189	0.301	2.176	0.301	1.408	0.370	18.387	0.912	13.608	0.074
	Prob	0.000	0.861	0.337	0.861	0.495	0.831	0.000	0.634	0.001	0.964
	Lag	2	2	2	2	2	2	2	2	2	2
		Y→G		No causality		No causality		Y→G		Y→G	
	Chi square	18.435	1.259	3.333	1.259	2.346	2.023	8.614	3.022	15.974	24.433
	Prob	0.000	0.739	0.343	0.739	0.504	0.568	0.035	0.388	0.007	0.000
	Lag	3	3	3	3	3	3	3	3	5	5
		Y→G		No causality		No causality		Y→G		Bi-directional	
10	Chi square	39.510	1.100	15.197	1.100	15.293	1.029	36.183	0.832	33.692	1.501
	Prob	0.000	0.577	0.001	0.577	0.001	0.598	0.000	0.660	0.000	0.472
	Lag	2	2	2	2	2	2	2	2	2	2
		Y→G		Y→G		Y→G		Y→G		Y→G	
	Chi square	39.449	6.960	11.257	6.960	11.097	7.227	45.733	1.348	32.579	9.442
	Prob	0.000	0.138	0.024	0.138	0.026	0.124	0.000	0.718	0.000	0.051
	Lag	4	4	4	4	4	4	3	3	4	4
		Y→G		Y→G		Y→G		Y→G		Bi-directional	
11	Chi square	29.725	3.015	15.849	3.015	14.745	6.003	8.567	17.115	22.147	8.894
	Prob	0.000	0.389	0.001	0.389	0.002	0.111	0.036	0.001	0.000	0.031
	Lag	3	3	3	3	3	3	3	3	3	3
		Y→G		Y→G		Y→G		Bi-directional		Bi-directional	
	Chi square	25.658	2.969	11.814	2.969	10.254	7.086			17.014	8.491
	Prob	0.000	0.563	0.019	0.563	0.036	0.131			0.002	0.075
	Lag	4	4	4	4	4	4			4	4
		Y→G		Y→G		Y→G				Bi-directional	

Note. The expenditure area numbers represent respectively: total expenditure (1), economic affairs (2), defense (3), education (4), health (5), public order and safety (6), social protection (7), general public services (8), environmental protection (9), recreation, culture, religion (10), and housing and community amenities (11).

Table 4. Dumitrescu-Hurlin panel causality test results.

Expenditure area	Specification 1 Peacock-Wiseman		Specification 2 Mann		Specification 3 Musgrave		Specification 4 Gupta		Specification 5 Goffman								
	D(G)	D(Y)	D(G/Y)	D(Y)	D(Y)	D(G/Y)	D(G/Y)	D(Y/P)	D(Y/P)	D(G/P)	D(Y/P)	D(Y/P)	D(G)	D(Y/P)	D(Y/P)	D(G)	
1	W-bar	3.241	0.968	0.895	0.968	0.990	0.866	3.248	0.744	3.319	0.945						
	Z-bar	8.386	-0.120	-0.392	-0.120	-0.036	-0.501	8.412	-0.957	8.677	-0.207						
	p-value	0.000	0.905	0.695	0.905	0.972	0.617	0.000	0.339	0.000	0.836						
	Z-bar tilde	6.490	-0.457	-0.680	-0.457	-0.389	-0.768	6.511	-1.141	6.727	-0.528						
	p-value	0.000	0.648	0.497	0.648	0.698	0.442	0.000	0.254	0.000	0.597						
		Y→G		No causality		No causality		Y→G		Y→G							
2	W-bar	2.233	1.586	1.575	1.586	1.601	1.476	2.327	1.312	2.232	1.531						
	Z-bar	4.614	2.191	2.153	2.191	2.247	1.779	4.963	1.167	4.608	1.986						
	p-value	0.000	0.029	0.031	0.029	0.025	0.075	0.000	0.243	0.000	0.047						
	Z-bar tilde	3.409	1.430	1.399	1.430	1.476	1.094	3.694	0.593	3.404	1.262						
	p-value	0.001	0.153	0.162	0.153	0.140	0.274	0.000	0.553	0.001	0.207						
		Y→G		No causality		No causality		Y→G		Y→G							
3	W-bar	1.696	1.916	1.393	1.916	1.430	1.881	2.379	1.744	1.657	1.883						
	Z-bar	2.605	3.426	1.472	3.426	1.609	3.295	5.161	2.783	2.456	3.303						
	p-value	0.009	0.001	0.141	0.001	0.108	0.001	0.000	0.005	0.014	0.001						
	Z-bar tilde	1.768	2.439	0.842	2.439	0.955	2.332	3.856	1.914	1.647	2.338						
	p-value	0.077	0.015	0.400	0.015	0.340	0.020	0.000	0.056	0.100	0.019						
		Bi-directional		G→Y		G→Y		Bi-directional		Bi-directional							
4	W-bar	5.129	1.511	1.458	1.511	1.296	1.303	3.363	1.307	4.773	1.363						
	Z-bar	15.450	1.911	1.714	1.911	1.107	1.132	8.842	1.150	14.119	1.358						
	p-value	0.000	0.056	0.087	0.056	0.268	0.258	0.000	0.250	0.000	0.174						
	Z-bar tilde	12.260	1.201	1.040	1.202	0.545	0.565	6.863	0.580	11.172	0.749						
	p-value	0.000	0.230	0.298	0.230	0.586	0.572	0.000	0.562	0.000	0.454						
		Y→G		No causality		No causality		Y→G		Y→G							
5	W-bar	3.285	0.790	1.332	0.790	1.191	0.772	3.293	0.925	3.192	0.737						
	Z-bar	8.551	-0.785	1.242	-0.785	0.715	-0.853	8.580	-0.279	8.203	-0.983						
	p-value	0.000	0.432	0.214	0.432	0.474	0.394	0.000	0.780	0.000	0.326						
	Z-bar tilde	6.625	-1.001	0.655	-1.001	0.225	-1.056	6.648	-0.588	6.341	-1.162						
	p-value	0.000	0.317	0.513	0.317	0.822	0.291	0.000	0.557	0.000	0.245						
		Y→G		No causality		No causality		Y→G		Y→G							

Table 4. (Cont).

6	W-bar	2.680	1.362	1.119	1.362	1.105	1.556	2.414	1.373	2.523	1.577
	Z-bar	6.286	1.354	0.445	1.354	0.393	2.082	5.290	1.395	5.700	2.158
	p-value	0.000	0.176	0.656	0.176	0.695	0.037	0.000	0.163	0.000	0.031
	Z-bar tilde	4.775	0.746	0.004	0.746	-0.039	1.341	3.961	0.780	4.296	1.403
	p-value	0.000	0.455	0.997	0.456	0.969	0.180	0.000	0.436	0.000	0.161
			Y→G		No causality		No causality		Y→G		Y→G
7	W-bar	3.252	1.590	2.376	1.590	2.607	1.761	1.954	1.969	3.111	1.621
	Z-bar	8.428	2.206	5.150	2.206	6.013	2.848	3.570	3.627	7.900	2.324
	p-value	0.000	0.027	0.000	0.027	0.000	0.004	0.000	0.000	0.000	0.020
	Z-bar tilde	6.524	1.443	3.847	1.443	4.552	1.966	2.556	2.603	6.093	1.538
	p-value	0.000	0.149	0.000	0.149	0.000	0.049	0.011	0.009	0.000	0.124
			Y→G		Y→G		Bi-directional		Bi-directional		Y→G
8	W-bar	1.950	2.596	0.913	2.596	0.991	2.371	1.404	2.147	2.027	2.555
	Z-bar	3.555	5.970	-0.326	5.970	-0.036	5.129	1.510	4.290	3.844	5.817
	p-value	0.000	0.000	0.744	0.000	0.972	0.000	0.131	0.000	0.000	0.000
	Z-bar tilde	2.544	4.517	-0.626	4.517	-0.389	3.830	0.874	3.145	2.780	4.391
	p-value	0.011	0.000	0.531	0.000	0.698	0.000	0.382	0.002	0.005	0.000
			Bi-directional		G→Y		G→Y		G→Y		Bi-directional
9	W-bar	2.119	0.750	1.352	0.750	1.212	0.769	2.614	1.402	1.909	0.789
	Z-bar	4.187	-0.936	1.316	-0.936	0.791	-0.863	6.040	1.505	3.403	-0.789
	p-value	0.000	0.349	0.188	0.349	0.429	0.388	0.000	0.132	0.001	0.430
	Z-bar tilde	3.060	-1.124	0.715	-1.124	0.287	-1.064	4.574	0.870	2.420	-1.004
	p-value	0.002	0.261	0.475	0.261	0.774	0.287	0.000	0.385	0.016	0.316
			Y→G		No causality		No causality		Y→G		Y→G
10	W-bar	3.285	1.021	1.740	1.021	1.667	1.032	3.060	0.661	3.279	1.072
	Z-bar	8.549	0.077	2.768	0.077	2.495	0.121	7.709	-1.270	8.529	0.269
	p-value	0.000	0.939	0.006	0.939	0.013	0.904	0.000	0.204	0.000	0.788
	Z-bar tilde	6.624	-0.297	1.901	-0.297	1.679	-0.261	5.937	-1.397	6.607	-0.140
	p-value	0.000	0.767	0.057	0.767	0.093	0.794	0.000	0.163	0.000	0.889
			Y→G		Y→G		Y→G		Y→G		Y→G
11	W-bar	1.408	1.073	1.013	1.073	0.948	1.088	1.413	1.207	1.347	1.060
	Z-bar	1.526	0.272	0.050	0.272	-0.196	0.329	1.546	0.776	1.299	0.224
	p-value	0.127	0.786	0.961	0.786	0.845	0.742	0.122	0.438	0.194	0.823
	Z-bar tilde	0.887	-0.138	-0.319	-0.138	-0.520	-0.091	0.903	0.274	0.702	-0.177
	p-value	0.375	0.891	0.750	0.891	0.603	0.927	0.366	0.784	0.483	0.860
			No causality		No causality		No causality		No causality		No causality

Note. As per Table 3.

Table 5. Bias-corrected Granger causality test results.

Expenditure area		Specification 1 Peacock-Wiseman		Specification 2 Mann		Specification 3 Musgrave				Specification 4 Gupta				Specification 5 Goffman					
		D(G)	D(Y)	D(G/Y)	D(Y)	D(Y)	D(G/Y)	D(G/Y)	D(Y/P)	D(Y/P)	D(G/Y)	D(G/P)	D(Y/P)	D(Y/P)	D(G/P)	D(G)	D(Y/P)	D(Y/P)	D(G)
1	HPJ	33.811	1.270	0.682	1.270	2.935	0.647	19.080	0.821	24.607	1.504								
	p-value	0.000	0.260	0.409	0.260	0.087	0.421	0.000	0.365	0.000	0.220								
	Estimator	0.385	-0.027	-0.066	-0.027	-0.136	-0.020	0.305	0.020	0.326	-0.030								
	p-value	0.000	0.260	0.409	0.260	0.087	0.421	0.000	0.365	0.000	0.220								
		Y→G			No causality		Y→G		Y→G		Y→G								
2	HPJ	0.137	0.482	6.632	0.482	8.524	0.220	0.523	0.063	0.000	0.335								
	p-value	0.711	0.487	0.010	0.487	0.004	0.639	0.470	0.801	0.986	0.563								
	Estimator	0.093	-0.004	-0.672	-0.004	-0.755	-0.003	-0.179	-0.001	0.005	-0.003								
	p-value	0.711	0.487	0.010	0.487	0.004	0.639	0.470	0.801	0.985	0.563								
		No causality			Y→G		Y→G		No causality		No causality								
3	HPJ	8.584	0.032	0.064	0.032	0.057	0.108	7.205	0.879	6.324	0.041								
	p-value	0.003	0.857	0.800	0.857	0.812	0.742	0.007	0.348	0.012	0.840								
	Estimator	0.559	0.002	0.049	0.002	-0.046	0.003	0.452	0.009	0.475	0.002								
	p-value	0.003	0.180	0.800	0.857	0.812	0.742	0.007	0.348	0.012	0.839								
		Y→G			No causality		No causality		Y→G		Y→G								
4	HPJ	76.201	0.429	2.222	0.429	0.912	1.000	40.331	7.728	70.472	0.334								
	p-value	0.000	0.513	0.136	0.513	0.340	0.317	0.000	0.005	0.000	0.563								
	Estimator	0.642	0.015	0.121	0.015	0.077	0.023	0.484	0.054	0.611	0.013								
	p-value	0.000	0.513	0.136	0.513	0.339	0.317	0.000	0.005	0.000	0.563								
		Y→G			No causality		No causality		Bi-directional		Y→G								
5	HPJ	52.275	20.878	5.965	20.878	3.099	19.524	35.574	18.975	47.383	17.096								
	p-value	0.000	0.000	0.015	0.000	0.078	0.000	0.000	0.000	0.000	0.000								
	Estimator	0.620	0.084	0.234	0.084	0.168	0.082	0.506	0.076	0.582	0.077								
	p-value	0.000	0.000	0.015	0.000	0.078	0.000	0.000	0.000	0.000	0.000								
		Bi-directional			Bi-directional		Bi-directional		Bi-directional		Bi-directional								

Table 5. (Cont).

6	HPJ	30.772	3.555	0.065	3.556	0.174	2.841	20.409	5.291	26.481	2.579
	p-value	0.000	0.059	0.799	0.059	0.676	0.092	0.000	0.021	0.000	0.108
	Estimator	0.731	0.025	0.033	0.025	-0.053	0.022	0.496	0.034	0.668	0.021
	p-value	0.000	0.059	0.799	0.059	0.676	0.092	0.000	0.021	0.000	0.108
		Bi-directional		G→Y		G→Y		Bi-directional		Y→G	
7	HPJ	39.533	0.229	0.490	0.229	0.112	0.161	9.266	4.865	30.563	0.038
	p-value	0.000	0.633	0.484	0.633	0.737	0.688	0.002	0.027	0.000	0.845
	Estimator	0.336	-0.015	0.063	-0.015	-0.031	0.013	0.198	0.057	0.293	-0.006
	p-value	0.000	0.632	0.484	0.632	0.737	0.688	0.002	0.027	0.000	0.845
		Y→G		No causality		No causality		Bi-directional		Y→G	
8	HPJ	15.101	0.801	0.008	0.801	0.110	0.811	10.690	0.288	13.703	0.934
	p-value	0.000	0.371	0.931	0.371	0.740	0.368	0.001	0.592	0.000	0.334
	Estimator	0.610	-0.009	-0.014	-0.009	-0.053	-0.009	0.523	-0.005	0.575	-0.010
	p-value	0.000	0.371	0.931	0.371	0.740	0.368	0.001	0.592	0.000	0.334
		Y→G		No causality		No causality		Y→G		Y→G	
9	HPJ	13.539	1.559	14.192	1.559	11.572	1.800	9.443	7.081	13.781	1.584
	p-value	0.000	0.212	0.000	0.212	0.001	0.180	0.002	0.008	0.000	0.208
	Estimator	0.996	0.008	0.976	0.008	0.877	0.008	0.875	0.016	0.995	0.008
	p-value	0.000	0.212	0.000	0.212	0.001	0.180	0.002	0.008	0.000	0.208
		Y→G		Y→G		Y→G		Bi-directional		Y→G	
10	HPJ	26.708	1.568	4.004	1.568	2.209	1.375	15.501	0.994	23.595	1.863
	p-value	0.000	0.211	0.045	0.211	0.137	0.241	0.000	0.319	0.000	0.172
	Estimator	0.743	-0.015	0.274	-0.015	0.202	-0.014	0.528	0.012	0.688	-0.016
	p-value	0.000	0.211	0.045	0.210	0.137	0.241	0.000	0.319	0.000	0.172
		Y→G		Y→G		No causality		Y→G		Y→G	
11	HPJ	20.952	0.145	6.709	0.145	4.520	0.109	7.875	5.655	17.461	0.164
	p-value	0.000	0.703	0.010	0.703	0.034	0.742	0.005	0.017	0.000	0.686
	Estimator	1.222	-0.002	0.677	-0.002	0.551	-0.002	1.007	-0.010	1.103	-0.002
	p-value	0.000	0.703	0.010	0.703	0.033	0.742	0.005	0.017	0.000	0.686
		Y→G		Y→G		Y→G		Bi-directional		Y→G	

Note. As per Table 3.

5. Discussion and conclusion

The paper examined alternative hypotheses that describe the relationship between public expenditure and output in a panel of 28 European economies over the 1995-2018 period using the panel data methods.

The respective variables were found to be cross-sectionally dependent (expected result, given the tight and extensive political and economic integration between European economies), and thus, the appropriate unit root and cointegration tests (Pesaran CADF and Westerlund) were used. The variables were characterised by a mix of $I(1)$ and $I(0)$ integration order, and the long-run equilibrium relationship between public expenditure and output was established in every specification and expenditure category. The strong evidence of cointegration is an indication of robust results, as Wagner's hypothesis presumes a strong equilibrium relationship between public expenditure and GDP at earlier stages of development, and a weaker relationship at later stages. This result is in line with other studies of Wagner's hypothesis for the advanced and industrialised economies in the post-WWII period (the study of G7 economies by Kolluri et al., 2000; of six European countries by Thornton, 1999; and of Italy by Magazzino, 2012).

The evidence of cointegration was somewhat weaker in the social protection expenditure category. While, as put by Shelton (2007), the population ageing in European countries has necessitated greater social security and protection expenditure and thus stronger relationship with GDP (per capita), the previously strong relationship between social expenditure and GDP in the well-established welfare states of Europe has weakened in recent decades, due to the structural change of the economy and the reforms of the social security and welfare systems that these economies have been undergoing in the recent decades (Carter, 2003; Kuckuck, 2014, p. 150). This likely affected the social protection category the most.

The causality running from output to public expenditure along Wagner's lines was identified in the majority of cases. The result was consistent across the spending categories. Given the robustness checks performed, it was also consistent across the periods as well as in the presence of cyclical dummies. The result supports the findings of previous research in European context, e.g. Magazzino (2011) and Lamartina and Zaghini (2011).

There were a number of instances, when there was no causality in either direction between GDP and public expenditure (14.0% of all cases in Toda-Yamamoto test, and 30.9% and 23.6% in Dumitrescu-Hurlin and Juodis-Karavias-Sarafidis tests). The absence of causality, however, was not a characteristic of a particular expenditure category, and thus can be attributed to the properties of the causality test or the specification of Wagner's hypothesis. The Keynesian hypothesis of sole influence of public expenditure on output and the absence of reverse causality did not find support (zero cases in Toda-Yamamoto test, and three and two cases in Dumitrescu-Hurlin and Juodis-Karavias-Sarafidis tests, respectively, in general public services and public order and safety expenditure categories). The missing causality in these two categories may be explained by the structure of the panel: the expansion of the public order and general administration activities alongside output growth is typical for developing economies with a nascent state apparatus, which is clearly not the case of European countries that

experienced such expansion in the 19th century. The bilateral causality was encountered in 37.0%, 10.9% and 20.0% of cases under the three tests. Overall, while bi-directional relationship between public expenditure and output was present, which to certain extent contradicts Wagner's hypothesis, the Keynesian hypothesis of sole unidirectional influence of public expenditure on output can be rejected (in this present study).

There are multiple policy implications of the results. Firstly, the strong evidence of cointegration between output and expenditure and the absence of government expenditure to output causality along Keynesian lines may imply limited effectiveness of the short-term spending cuts and rises in government expenditure, given that expenditure will tend to return to a long-run equilibrium level. In this context, as noted by Magazzino (2011) and Akitoby et al (2006), structural reforms would be needed to help achieve the social and economic objectives that fiscal policy is supposed to accomplish. Secondly, the strong output-to-expenditure causality along Wagner's lines and sustained public expenditure was observed despite the slowdown of European economic growth in recent decades. The factors that helped to maintain strong long-term relationships and causality along Wagner's lines included increased revenue collection of the governments, a decrease in the size of the shadow economy, and the growing demand for public expenditure in certain functional areas, e.g. increase in health spending due to new health threats, or the projected expansion of the defence expenditures due to uncertain geopolitical situation in Europe. Thirdly, as argued by Afonso et al (2005), the Wagner's relationship is necessarily weakened over time, due to the improvement of the quality of goods and services provided by the public sector, to the higher efficiency of the sector and slower increase (and perhaps) decrease in spending in the industrial economies. The analysis conducted in this paper for the sub-periods did not demonstrate any weakening of the 'output-spending' cointegration in later periods. Fourthly, as noted by Benavides et al (2013, pp. 71-72), the causality along Wagner's as opposed to Keynes' lines points to the limited role of political factors and political decision-making in the determination of economic outcomes. This empirical result would be contrary to the actual experience of the developed economies in the post-WWII period (expansion of public policies, growth of bureaucracy and the predominance of redistribution as opposed to efficiency objectives and logic in the public policies), and the premises of the public choice and new political economy theories (Buchanan, Tullock, 1977). Lastly, the identified weakness of Keynesian hypothesis may be attributed to the crowding out of private investment by public spending, a phenomenon that has been well documented in the literature (Ashauer, 1989, Erden, Holcombe, 2005). The relevant policy implication is the absence of any automatic positive effect of greater public spending on growth and productivity. In a related vein, the weakness of Keynesian causality may be explained by the weakening of the positive effects of public spending on total factor productivity that has been observed starting from the 1980s. With regard to both phenomena, Lachler and Aschauer (1998) stressed the importance of savings-financed increase in public spending and capital (as opposed to an increase in public consumption that merely brings greater public debt and higher current and future taxes) for the full manifestation of positive gains of fiscal effort.

Future research may modify the present model and setting by considering the cyclical components of government expenditure and output and alternative channels through which the expenditure is directed in the economy, by incorporating the effects of discretionary fiscal policy, and by investigating

Wagner's hypothesis at the level of sub-national finances (Kucukkale, Yamak, 2012; Inshauspe et al, 2020). Future empirical studies may consider the role of public enterprises (in addition to the level of spending), the role of government regulation and legislative activity as a proxy or measure of the government size, the breaks and non-linearities in the relationship (in line with Armey-Rahn hypothesis of optimal size of government). They may also continue to experiment with alternative functional forms of the relationship and alternative definitions of dependent and independent variables (Peacock, Scott, 2000).

The bulk of the research on Wagner's hypothesis has been on variables (e.g. public spending or the scope of legislative activity of the government) that are amenable to quantitative analysis. A more important issue is the nature of the government itself. As noted by Lamartina and Zaghini (2011), the pursuit of self- or vested-interests (alongside corruption and moral hazard behaviour) by the bureaucracy would render expansion of the government involvement in social and economic affairs undesirable, notwithstanding the growth of the economy and objective need for 'bigger' government. In a related vein, the growth of the spending on public services and administration that is motivated by the self-interest in the expansion of bureaucratic prominence and functions or proliferation of government programs due to lobbying efforts are examples of the growth of public spending that takes place irrespective of output or economic growth (Niskanen, 1971). Wagner's hypothesis is thus a partial explanation of the public sector growth.

References

- Abizadeh, S., Gray, J. (1985). Wagner's law: a pooled time series, cross-section analysis. *National Tax Journal* 38, 209-218. <https://doi.org/10.1086/NTJ41792010>
- Afonso, A., Alves, J. (2017). Reconsidering Wagner's law: evidence from the functions of the government. *Applied Economics Letters* 24(5), 346-350. <https://doi.org/10.1080/13504851.2016.1192267>
- Afonso, A., Jalles, J. (2014). Causality for the government budget and economic growth. *Lisboa School of Economics and Management Working Paper No. WP07/2014/DE/UECE*, pp. 1-5. <https://doi.org/10.2139/ssrn.2429909>
- Afonso, A., Schuknecht, L., Tanzi, V. (2005). Public sector efficiency: an international comparison. *Public Choice* 123, 321-347. <https://doi.org/10.1007/s11127-005-7165-2>
- Ahsan, M., Kwan, A., Sahni, B. (1996). Cointegration and Wagner's hypothesis: time series Evidence for Canada. *Applied Economics* 28(8), 1055-1058. <https://doi.org/10.1080/000368496328182>
- Akitoby, B., Clement, B., Gupta, S., Inchauste, G. (2006). Public spending, voracity, and Wagner's law in developing countries. *European Journal of Political Economy* 22(4), 908-924. <https://doi.org/10.1016/j.ejpoleco.2005.12.001>
- Alezina, A., Wacziarg, R. (1998). Openness, country size and government. *Journal of Public Economics* 69(3), 305-321. [https://doi.org/10.1016/S0047-2727\(98\)00010-3](https://doi.org/10.1016/S0047-2727(98)00010-3)
- Ansari, M., Gordon, D., Akuamoah, C. (1997). Keynes versus Wagner: public expenditure and national income for three African countries. *Applied Economics* 29(4), 543-550. <https://doi.org/10.1080/000368497327038>
- Apergis, N., Christou, C., Hassapis, C. (2013). Convergence in public expenditures across EU countries: evidence from club convergence. *Economic and Finance Research: An Open Access Journal* 1(1), 45-59. <https://doi.org/10.1080/21649480.2013.862735>
- Aschauer, D. (1989). Public investment and productivity growth in the Group of Seven. *Economic Perspectives*, Federal Reserve Bank of Chicago 13, 17-25.

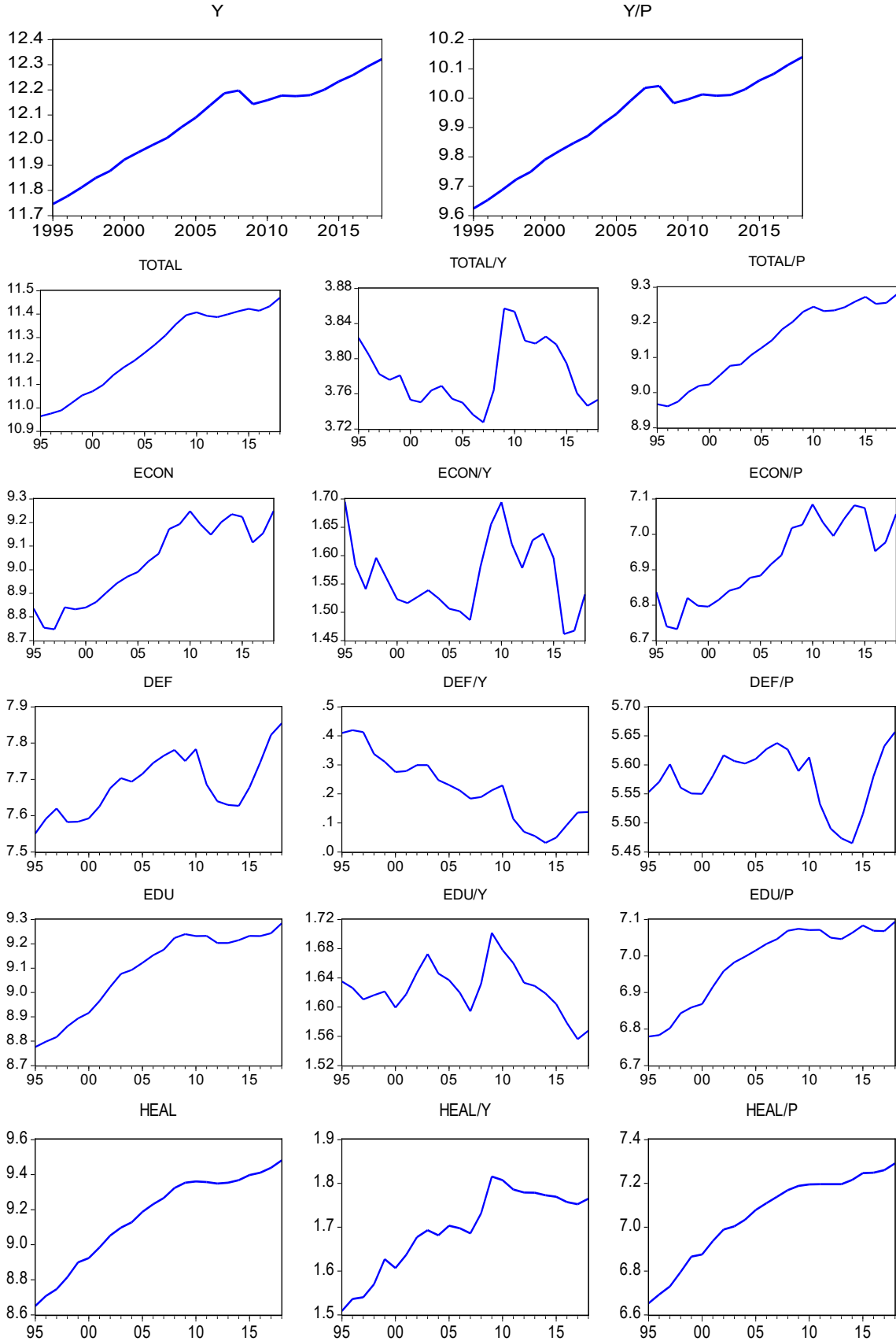
- Ayad, H., Belmokaddem, M. (2017). Financial development, trade openness and economic growth in MENA countries: TYDL panel causality approach. *Theoretical and Applied Economics* 24(1), 233-246.
- Azcona, N. (2022). Business cycle co-movement in Europe: trade, industry composition and the single currency. *Open Economies Review* 33(1), 121-139. <https://doi.org/10.1007/s11079-021-09625-7>
- Babatunde, M. (2011). A bound testing analysis of Wagner's law in Nigeria 1970-2006. *Applied Economics* 43(21), 2843-2850. <https://doi.org/10.1080/00036840903425012>
- Bairam, E. (1992). Variable elasticity and Wagner's law. *Public Finance* 3, 491-495.
- Barro, R. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics* 106(2), 407-443. <https://doi.org/10.2307/2937943>
- Baumol, W. (1967). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *American Economic Review* 57 (3), 415-426.
- Benavides, D., Venegas-Martinez, F., Santiago, V. (2013). La ley de Wagner versus la hipotesis keynesiana: el case de Mexico, 1950-2009. *Investigacion Economica LXXII(283)*, 69-98. [https://doi.org/10.1016/S0185-1667\(13\)72587-6](https://doi.org/10.1016/S0185-1667(13)72587-6)
- Biswal, B., Dhawan, U., Lee, H.-Y. (1999). Testing Wagner versus Keynes using disaggregated public expenditure data for Canada. *Applied Economics* 31(10), 1283-1291. <https://doi.org/10.1080/000368499323490>
- Bohl, M. (1996). Some international evidence on Wagner's law. *Public Finance* 51(2), 185-200.
- Buchanan, J., Tullock, G. (1977). The expanding public sector: Wagner squared. *Public Choice* 31, 147-150. <https://doi.org/10.1007/BF01718982>
- Burney, N. (2002). Wagner's hypothesis: evidence from Kuwait using cointegration tests. *Applied Economics* 34(1), 49-57. <https://doi.org/10.1080/00036840010027540>
- Caporale, G., Pittis, N. (1999). Unit root testing using covariates: some theory and evidence. *Oxford Bulletin of Economics and Statistics* 61(4), 583-595. <https://doi.org/10.1111/1468-0084.00145>
- Carter, M. (2003). Three essays on empirical patterns in functional distribution for a cross section of countries: evidence from the Extended Penn World Tables, 1963-1996. New School University, Ph.D. Dissertation.
- Chang, T., Liu, W., Caudill, S. (2004). A re-examination of Wagner's law for ten countries based on cointegration and error-correction modelling techniques. *Applied Financial Economics* 14(8), 577-589. <https://doi.org/10.1080/0960310042000233872>
- Courakis, A., Moura-Roque, F., Tridimas, G. (1993). Public expenditure growth in Greece and Portugal: Wagner's law and beyond. *Applied Economics* 25(1), 125-134. <https://doi.org/10.1080/00036849300000121>
- Cuddington, J., Ludema, R., Jayasuriya, S. (2002). *Prebisch-Singer Redux*. Washington: US International Trade Commission.
- Dao, M. (1995). Determinants of government expenditures: new evidence from disaggregative data. *Oxford Bulletin of Economics and Statistics* 57(1), 67-76. <https://doi.org/10.1111/j.1468-0084.1995.tb00027.x>
- Diamond, J. (1977). Wagner's "Law" and the developing countries. *The Developing Economies* 15(1), 37-59. <https://doi.org/10.1111/j.1746-1049.1977.tb00369.x>
- Dollery, B., Singh, S. (2000). Explaining the real size of government in Australia: an application of the Ferris and West model. *Economic Analysis and Policy* 30(2), 157-173. [https://doi.org/10.1016/S0313-5926\(00\)50018-6](https://doi.org/10.1016/S0313-5926(00)50018-6)
- Dumitrescu, E.-I., Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling* 29(4), 1450-1460. <https://doi.org/10.1016/j.econmod.2012.02.014>
- Erden, L., Holcombe, R. (2005). The effects of public investment on private investment in developing economies. *Public Finance Review* 33(5), 575-602. <https://doi.org/10.1177/1091142105277627>
- Gandhi, V. P. (1971). Wagner's law of public expenditure: do recent cross-section studies confirm it? *Public Finance* 26, 44-56.
- Ghazy, N., Ghoneim, H., Pappas, D. (2021). The validity of Wagner's law in Egypt from 1960-2018. *Review of Economics and Political Science*, 6(2), 98-116. <https://doi.org/10.1108/REPS-01-20200004>
- Goffman, J. (1968). On the empirical testing of Wagner's law: a technical note. *Public Finance* 3(3), 359-364.
- Granger, C.W.J. (1988). Aggregation of time series variables - a survey. Federal Reserve Bank of Minneapolis, Discussion Paper No. 1, pp. 1-26. <https://doi.org/10.21034/dp.1>
- Gupta, S. (1967). Public expenditure and economic growth: a time series analysis. *Public Finance* 22(4), 423-461.

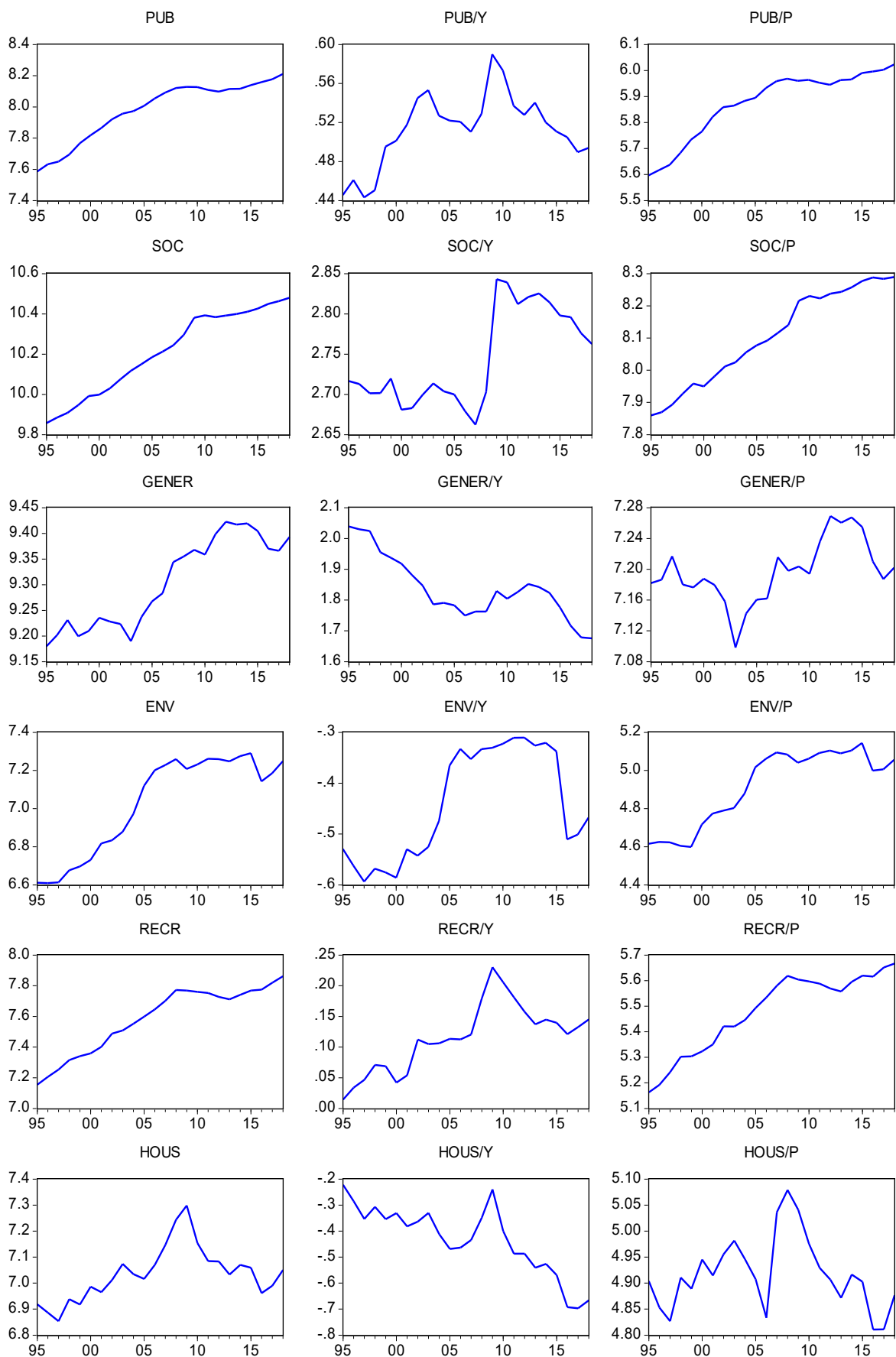
- Huang, C.-J. (2006). Government expenditures in China and Taiwan: do they follow Wagner's law? *Journal of Economic Development* 31(2), 139-148.
- Inchauspe, J., Kobir, M., MacDonald, G. (2020). Wagner's law and the dynamics of government spending in Indonesia. *Bulletin of Indonesian Economic Studies* 58(1), 1-22. <https://doi.org/10.1080/00074918.2020.1811837>
- Islam, A. (2001). Wagner's law revisited: cointegration and exogeneity tests for USA. *Applied Economics* 8, 509-515. <https://doi.org/10.1080/13504850010018743>
- Iyare, S., Lorde, T. (2004). Co-integration, causality and Wagner's law: tests for selected Caribbean economies. *Applied Economics Letters* 11(13), 815-825. <https://doi.org/10.1080/1350485042000254881>
- Jaen Garcia, M. (2004). La ley de Wagner: un analisis sintetico. Instituto de Estudios Fiscales, Working Paper No. 6/04, pp. 1-59.
- Juodis, A., Karavias, Y., Sarafidis, V. (2021). A homogeneous approach to testing for Granger non-causality in heterogeneous panels. *Empirical Economics* 60, 93-112. <https://doi.org/10.1007/s00181-020-01970-9>
- Karagianni, S., Pempetzoglou, M. (2009). Evidence for non-linear causality between public spending and income in the European Union countries. *Journal of Applied Business Research* 25, 69-82. <https://doi.org/10.19030/jabr.v25i1.1049>
- Koester, G., Priesmeier, C. (2013). Does Wagner's law ruin the sustainability of German public finances? *FinanzArchiv / Public Finance* 69(3), 256-288. <https://doi.org/10.1628/001522113X671128>
- Kolluri, B., Panik, M., Wahab, M. (2000). Government expenditure and economic growth: evidence from G7 countries. *Applied Economics* 32(8), 1059-1068. <https://doi.org/10.1080/000368400322110>
- Kuckuck, J. (2014). Testing Wagner's law at different stages of economic development. *FinanzArchiv / Public Finance Analysis* 70(1), 128-168. <https://doi.org/10.1628/001522114X679183>
- Kucukkale, Y., Yamak, R. (2012). Cointegration, causality and Wagner's law with disaggregated data: evidence from Turkey, 1968-2004. MPRA Paper No. 36894.
- Lachler, U., Aschauer, D. (1998). Public investment and economic growth in Mexico. World Bank, Policy Research Working Paper No. 1964.
- Lamartina, S., Zaghini, A. (2011). Increasing public expenditure: Wagner's law in OECD countries. *German Economic Review* 12(2), 149-164. <https://doi.org/10.1111/j.1468-0475.2010.00517.x>
- Legrenzi, G. (2000). The long-run relationship between public sector size and economic growth: income-elasticity and causality of the Italian general government expenditure (1861-1998). *Rivista Internazionale di Scienze Economiche e Commerciali* 47(3), 415-437.
- Lin, C.-A. (1995). More evidence on Wagner's law for Mexico. *Public Finance* 50(2), 267-277.
- Loizides, J., Vamvoukas, G. (2005). Government expenditure and economic growth: evidence from trivariate causality testing. *Journal of Applied Economics* 8(1), 125-152. <https://doi.org/10.1080/15140326.2005.12040621>
- Lopez, L., Weber, S. (2017). Testing for Granger causality in panel data. University of Neuchatel, Institute for Economic Research Working Paper No. 17-03, pp. 1-13. <https://doi.org/10.1177/1536867X1701700412>
- Magazzino, C. (2011). Disaggregated public spending, GDP and money supply: evidence for Italy. *European Journal of Economics, Finance and Administrative Sciences* 41, 118-131.
- Magazzino, C. (2012). Wagner versus Keynes: public spending and national income in Italy. *Journal of Policy Modeling* 34(6), 890-905. <https://doi.org/10.1016/j.jpolmod.2012.05.012>
- Mann, A. (1980). Wagner's Law: an econometric test for Mexico, 1925-1970. *National Tax Journal* 33, 189-201. <https://doi.org/10.1086/NTJ41862301>
- Mueller, D. (1987). The growth of government: a public choice perspective. *IMF Staff Papers* 34(1), 115-149. <https://doi.org/10.2307/3867026>
- Musgrave, R. (1969). *Fiscal systems*. New Haven: Yale University Press.
- Narayan, P., Nielsen, I., Smyth, R. (2008). Panel data, cointegration, causality and Wagner's law: empirical evidence from Chinese provinces. *China Economic Review* 19(2), 297-307. <https://doi.org/10.1016/j.chieco.2006.11.004>
- Narayan, S., Rath, B., Narayan, P. (2012). Evidence of Wagner's law from Indian state. *Economic Modelling* 29(5), 1548-1557. <https://doi.org/10.1016/j.econmod.2012.05.004>
- Niskanen, W. (1971). *Bureaucracy and representative government*. Chicago: Aldine-Atherton.
- North, D., Wallis, J. (1982). Longer-run determinants of the economic role of the US government. *American Economic Review (Papers and Proceedings)* 72(2), 336-340.
- Organisation for Economic Cooperation and Development / OECD (2011). *Government at a glance*, 2011. Paris: OECD Publishing.

- Oxley, L. (1994). Cointegration, causality and Wagner's law: a test for Britain 1870-1913. *Scottish Journal of Political Economy* 41(3), 286-298. <https://doi.org/10.1111/j.1467-9485.1994.tb01127.x>
- Peacock, A., Wiseman, J. (1979). Approaches to the analysis of government expenditure growth. *Public Finance Quarterly* 7(1), 3-23. <https://doi.org/10.1177/109114217900700101>
- Peacock, A., Scott, A. (2000). The curious attraction of Wagner's law. *Public Choice* 102, 1-17. <https://doi.org/10.1023/A:1005032817804>
- Persyn, D., Westerlund, J. (2008). Error-correction-based cointegration tests for panel data. *The Stata Journal* 8(2), 232-241. <https://doi.org/10.1177/1536867X0800800205>
- Rajaguru, G. (2002). The distortionary effects of temporal aggregation on Granger causality. *Bond University Business School Publications*, pp. 1-37.
- Ray, P., Pal, P. (2022). Fiscal stimulus and the ghost of Keynes: an evolutionary chronicle. In N. Yoshino, R. Paramanik, & A. Kumar (Eds.), *Studies in international economics and finance*. Singapore: Springer. https://doi.org/10.1007/978-981-16-7062-6_4
- Sedrakyan, G., Varela-Candamio, L. (2019). Wagner's law vs. Keynes' hypothesis in very different countries (Armenia and Spain). *Journal of Policy Modeling* 41(4), 747-762. <https://doi.org/10.1016/j.jpolmod.2019.02.011>
- Shelton, C. (2007). The size and composition of government expenditure. *Journal of Public Economics* 91(11-12), 2230-2260. <https://doi.org/10.1016/j.jpubeco.2007.01.003>
- Sideris, D. (2007). Wagner's law in 19th century Greece: a cointegration and causality analysis. *Bank of Greece Working Paper No. 64*, pp. 1-18. <https://doi.org/10.2139/ssrn.4164000>
- Sinha, D. (2007). Does the Wagner's law hold for Thailand? A time series study. *MPRA Paper No. 2560*.
- Tang, T.C. (2001). Testing the relationship between government expenditure and national income in Malaysia, *Analisis* 8(1/2), 37-51.
- Thornton, J. (1999). Cointegration, causality and Wagner Law in 19th century Europe. *Applied Economics Letters* 6(7), 413-416. <https://doi.org/10.1080/135048599352916>
- Timm, H. (1961). Das gesetz der wachsenden staatsausgaben. *Finanzarchiv NF* 21, 201-247.
- Tobin, D. (2005). Economic liberalization, the changing role of the state and "Wagner's law": China's development experience since 1978. *World Development* 33(5), 729-743. <https://doi.org/10.1016/j.worlddev.2004.12.001>
- Toda, H., Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66(1-2), 225-250. [https://doi.org/10.1016/0304-4076\(94\)016168](https://doi.org/10.1016/0304-4076(94)016168)
- Wagner, A. (1912). *Les fondements de l'economie politique*. Paris: Girard & Briere.
- Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69(6), 709-748. <https://doi.org/10.1111/j.1468-0084.2007.00477.x>
- Xiao, J., Juodis, A., Karavias, Y., Sarafidis, V. (2021). Improved tests for Granger non-causality in panel data. *Munich: MPRA Paper No. 107180*, pp. 1-11.
- Yay, T., Tastan, H. (2009). Growth of public expenditures in Turkey during the 1950-2004 period: an econometric analysis. *Romanian Journal of Economic Forecasting* 4, 101-118.
- Ziramba, E. (2009). Wagner's law: an econometric test for South Africa, 1960-2006. *South African Journal of Economics* 76(4), 596-606. <https://doi.org/10.1111/j.1813-6982.2008.00218.x>

Appendix

Figure 1. Dynamics of output and expenditure categories





Note. The series are expressed in natural logarithms.

Table 6. Cross-sectional results.

Expenditure area	Coefficient	Specification 1 Peacock-Wiseman		Specification 2 Mann		Specification 3 Musgrave		Specification 4 Gupta		Specification 5 Goffman	
		Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Total	C	-1.319	0.000	3.286	0.000	3.106	0.000	-1.310	0.001	1.066	0.767
	Cross-sect.	1.041	0.000	0.041	0.020	0.068	0.058	1.053	0.000	1.027	0.008
Economic affairs	C	-2.446	0.000	2.159	0.000	2.525	0.000	-1.855	0.001	0.485	0.886
	Cross-sect.	0.953	0.000	-0.047	0.074	-0.094	0.073	0.887	0.000	0.865	0.016
Defense	C	-4.801	0.000	-0.196	0.773	2.109	0.054	-2.338	0.043	0.069	0.986
	Cross-sect.	1.036	0.000	0.036	0.515	-0.188	0.084	0.799	0.000	0.771	0.061
Education	C	-2.778	0.000	1.827	0.000	0.952	0.046	-3.474	0.000	-1.088	0.745
	Cross-sect.	0.984	0.000	-0.016	0.491	0.068	0.148	1.054	0.000	1.026	0.005
Health	C	-3.937	0.000	0.669	0.187	0.678	0.432	-3.673	0.000	-1.362	0.722
	Cross-sect.	1.086	0.000	0.086	0.044	0.103	0.237	1.082	0.000	1.062	0.010
Public order and safety	C	-3.611	0.000	0.994	0.024	3.261	0.000	-1.230	0.011	1.221	0.735
	Cross-sect.	0.961	0.961	-0.039	0.262	-0.275	0.000	0.717	0.000	0.683	0.068
Social protection	C	-2.969	0.000	1.636	0.000	0.927	0.080	-3.473	0.000	-1.113	0.764
	Cross-sect.	1.092	0.000	0.092	0.001	0.183	0.001	1.167	0.000	1.142	0.005
General public services	C	-3.447	0.000	1.158	0.015	1.231	0.115	-3.285	0.000	-0.809	0.828
	Cross-sect.	1.058	0.000	0.058	0.128	0.063	0.414	1.057	0.000	1.021	0.011
Environmental protection	C	-5.816	0.000	-1.211	0.054	-0.996	0.333	-5.287	0.000	-3.035	0.431
	Cross-sect.	1.067	0.000	0.067	0.183	0.060	0.556	1.033	0.000	1.019	0.013
Recreation, culture, religion	C	-3.998	0.000	0.607	0.220	0.013	0.987	-4.303	0.000	-2.027	0.548
	Cross-sect.	0.960	0.000	-0.040	0.327	0.012	0.886	0.987	0.000	0.970	0.008
Housing and community amenities	C	-4.587	0.000	0.018	0.982	1.765	0.172	-2.586	0.059	-0.274	0.942
	Cross-sect.	0.968	0.000	-0.032	0.633	-0.215	0.100	0.762	0.000	0.744	0.058

Table 7. Toda-Yamamoto panel causality test with cyclical dummy results.

Expenditure area		Specification 1 Peacock-Wiseman			Specification 2 Mann			Specification 3 Musgrave				Specification 4 Gupta				Specification 5 Goffman			
		D(G)	D(Y)	D(Y)	D(G/Y)	D(Y)	D(Y)	D(G/Y)	D(G/Y)	D(Y/P)	D(Y/P)	D(G/Y)	D(G/P)	D(Y/P)	D(Y/P)	D(G/P)	D(G)	D(Y/P)	D(Y/P)
1	Lag	4	4	4	4	4	4	4	4	4	4	2	2	4	4	4	4	4	4
		Bi-directional			Bi-directional			Bi-directional				Y→G				Bi-directional			
2	Lag	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5
		Bi-directional			Bi-directional			Bi-directional				Bi-directional				Bi-directional			
3	Lag	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
		Bi-directional			Bi-directional			Bi-directional				Bi-directional				Bi-directional			
3	Lag	3	3	3	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3
		Y→G			No causality			No causality				Y→G				Y→G			
4	Lag	7	7	7	7	7	7	7	7	7	7	3	3	8	8	8	8	8	8
		Y→G			Y→G			Y→G				Y→G				Y→G			
4	Lag	4	4	4	4	4	4	4	4	4	4	3	3	4	4	4	4	4	4
		Y→G			Y→G			Y→G				Y→G				Y→G			
5	Lag	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
		Y→G			Y→G			Bi-directional				Bi-directional				Bi-directional			
6	Lag	8	8	8	8	8	8	6	6	6	6	8	8	6	6	6	6	6	6
		Bi-directional			Bi-directional			Bi-directional				Bi-directional				Y→G			
6	Lag	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
		Y→G			No causality			Bi-directional				Y→G				Bi-directional			
7	Lag	4	4	4	4	4	4	4	4	4	4	5	5	4	4	4	4	4	4
		Y→G			Y→G			No causality				Y→G				Y→G			
7	Lag	3	3	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5
		Bi-directional			Bi-directional			Bi-directional				Y→G				Bi-directional			
8	Lag	5	5	5	5	5	5	3	3	3	3	3	3	3	3	3	3	3	3
		Bi-directional			Bi-directional			Bi-directional				Y→G				Bi-directional			
8	Lag	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
		No causality			Y→G			Y→G				No causality				No causality			
8	Lag	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
		No causality			Y→G			Y→G				No causality				No causality			

Table 7. (Cont).

9	Lag	2	2	2	2	2	2	2	2	2	2
		Y→G		No causality		No causality		Y→G		Y→G	
	Lag	3	3	3	3	3	3	3	3	5	5
		Y→G		No causality		No causality		Y→G		Bi-directional	
10	Lag	2	2	2	2	2	2	2	2	2	2
		Y→G		Y→G		No causality		Y→G		Y→G	
	Lag	4	4	4	4	4	4	3	3	4	4
		Y→G		Y→G		Bi-directional		Y→G		Bi-directional	
11	Lag	3	3	3	3	3	3	3	3	3	3
		Y→G		Y→G		Bi-directional		Bi-directional		Bi-directional	
	Lag	4	4	4	4	4	4			4	4
		Y→G		Y→G		Bi-directional				Bi-directional	

Note. As per Table 3. The causality patterns that are different from the Toda-Yamamoto procedure with no cyclical dummy (Table 3) are highlighted in bold.

Table 8. Toda-Yamamoto panel causality test with positive and negative output components (whole period and sub-periods)

Expenditure area		Whole period		2009-2018		1995-2008	
		Specification 1		Specification 1		Specification 1	
		Peacock-Wiseman		Peacock-Wiseman		Peacock-Wiseman	
		D(G)	D(Y)	D(G)	D(Y)	D(G)	D(Y)
1	Lag	4	4	4	4	4	4
	Y+	Bi-directional		Bi-directional		Y→G	
	Y-	Bi-directional		Bi-directional		G→Y	
	Lag	5	5				
	Y+	Bi-directional					
	Y-	Bi-directional					
2	Lag	4	4	4	4	4	4
	Y+	Bi-directional		Bi-directional		Y→G	
	Y-	Y→G		Y→G		No causality	
3	Lag	3	3	3	3	3	3
	Y+	Y→G		Y→G		No causality	
	Y-	Y→G		Y→G		No causality	
	Lag	7	7				
	Y+	Bi-directional					
	Y-	Y→G					
4	Lag	4	4	4	4	4	4
	Y+	Y→G		Y→G		Bi-directional	
	Y-	Y→G		Bi-directional		Y→G	
5	Lag	4	4	4	4	4	4
	Y+	Y→G		Y→G		Y→G	
	Y-	Bi-directional		Bi-directional		Y→G	
	Lag	8	8				
	Y+	Y→G					
	Y-	Bi-directional					
6	Lag	2	2	2	2	2	2
	Y+	Y→G		Y→G		No causality	
	Y-	Y→G		Y→G		Bi-directional	
	Lag	4	4				
	Y+	Y→G					
	Y-	Y→G					
7	Lag	3	3	3	3	3	3
	Y+	Y→G		Bi-directional		Y→G	
	Y-	Bi-directional		Bi-directional		No causality	
	Lag	5	5				
	Y+	Bi-directional					
	Y-	Y→G					
8	Lag	2	2	2	2	2	2
	Y+	No causality		No causality		No causality	
	Y-	No causality		Y→G		G→Y	
	Lag	3	3				
	Y+	Y→G					
	Y-	No causality					

Table 8. (Cont).

9	Lag	2	2	2	2	2	2
	Y+	Y→G		No causality		Y→G	
	Y-	No causality		No causality		No causality	
	Lag	3	3				
	Y+	Y→G					
	Y-	No causality					
10	Lag	2	2	2	2	2	2
	Y+	Y→G		No causality		Y→G	
	Y-	Y→G		Y→G		Bi-directional	
	Lag	4	4				
	Y+	Y→G					
	Y-	Y→G					
11	Lag	3	3	3	3	3	3
	Y+	Y→G		Y→G		Y→G	
	Y-	Y→G		Y→G		Bi-directional	
	Lag	4	4				
	Y+	Y→G					
	Y-	Y→G					