Int Tax Public Finance (2017) 24:652–677
DOI 10.1007/s10797-017-9458-z

Wagner's law, fiscal discipline, and intergovernmental transfer: empirical evidence at the US and German state levels

Yoshito Funashima[1] · **Kazuki Hiraga**[2]

Published online: 30 May 2017

Abstract Does fiscal discipline restrain the government from increasing its budget size? To answer this question, this paper investigates whether Wagner's law is satisfied for two types of states: US states, in which fiscal sovereignty is established, and German states, in which fiscal transfer dependence is high and budget constraints are softened. In US states, we demonstrate that Wagner's law is validated, while some of the balanced budget requirements weaken the validity of the law. In German states, we find an "inverse" law, especially after the bailouts of Bremen and Saarland. The "inverse" law is a new channel of growth in government size and means that soft budget constraints cause significant negative correlation between government size and output. These results are robust regardless of whether intergovernmental fiscal transfers are taken into account, while they quantitatively change the validity of the law. Our findings imply that the characteristics of fiscal discipline are the prime determinants of the channel and degree of growth in government size.

Keywords Wagner's law · Fiscal discipline · State government · Balanced budget requirements · Soft budget constraints

JEL Classification H72 · C23

✉ Yoshito Funashima
funashima@mail.tohoku-gakuin.ac.jp

Kazuki Hiraga
khiraga581470@gmail.com

1 Faculty of Economics, Tohoku Gakuin University, 1-3-1 Tsuchitoi, Aoba-ku, Sendai, Miyagi 980-8511, Japan

2 School of Political Science and Economics, Tokai University, 4-1-1 Kitakaname, Hiratsuka 259-1292, Japan

1 Introduction

A number of industrialized economies have experienced substantial public sector growth over the last few centuries. The most prominent theory purporting to explain this long-run growth was proposed by the German political economist Adolph Wagner more than a century ago. His view is now commonly known as Wagner's law of increasing state activity, in which there is a positive relationship between economic development and government size.[1] Many researchers have been interested in the validity of the law, which is a crucial factor for fiscal rigidity and the government debt problem. If the law is supported, then government size (government expenditure as a share of output) increases as the economy (real output per capita) grows; as a result, the government is likely to be forced to abandon a flexible fiscal policy or to borrow excessively. With this background, over the past century, Wagner's law has been intensively tested for many countries, states, and periods. However, earlier empirical works find mixed results for the validity of the law, and there is no general consensus on the law among researchers.[2]

Our research question asks why the literature investigating Wagner's law does not have consensus and fails to obtain a reasonable estimation of the law. An answer for this question might be related to the fact that, thus far, no empirical studies have underscored the relevance of fiscal discipline as a precondition for Wagner's law. In general, it is likely that fiscal discipline can be considered a key element of growth in government expenditure. For example, the balanced budget requirements may contribute to the restraint of government expenditure, weakening the validity of the law.

In a contrasting situation, there is a possibility that lack of fiscal discipline is also relevant to Wagner's law. To be precise, soft budget constraints might allow the governments to expand expenditures without limits (e.g., Oates 2005; Weingast 2009). In fact, as shown in Rodden (2005), the ossification of the bailout expectations of the German state governments is related to larger deficits, since they are less willing to cut politically painful expenditures in response to negative shocks such that revenue is below expectation. Under the softening of budget constraints, it is conceivable that, even when facing low growth and tax revenue shortfall, governments rely heavily on debt and expand expenditures; as a result, the share of government expenditure in output (i.e., government size) increases. In this case, it is not economic development, but rather economic stagnation that causes growth in government size. That is, a negative correlation between government size and economic development, an "inverse" law, can arise under soft budget constraints (strong bailout expectations).

Motivated by the above hypotheses, in this paper, we test whether the validity of Wagner's law depends on the degree of fiscal discipline. To accomplish this, we

[1] In Narayan et al. (2012), Wagner's law is considered public expenditure behavior, such that "as real income increases, over the long-run, the share of public expenditure relative to national income rises" (p. 1548). While, as Peacock and Scott (2000) suggest, the law does not build on an explicit theoretical framework by which cause and effect are clearly explained, Rowley and Tollison (1994) document that the law is in accord with the idea of comparative advantage.

[2] See, for example, Kolluri et al. (2000) for support for the law, Chang (2002) for qualified support, and Shelton (2007) for no support. Although the literature on Wagner's law is too voluminous to survey here, an extensive survey of the literature can be found in Durevall and Henrekson (2011, p. 720–721).

focus on two federal states: the US and Germany. This is because fiscal discipline at the state level is quite different between these two countries, as stated in Bordo et al. (2013) and Potrafke and Reischmann (2015). In US states, in addition to a high degree of expenditure decentralization, tax bases and rates are determined with discretionary powers, and fiscal sovereignty is established. On the other hand, in German states, although the expenditure is restrained by and almost all tax bases and rates are determined by federal law, the state governments have full autonomy in borrowing. More importantly, the US and German federal governments officially follow the no-bailout rule; however, in Germany, the credibility of the commitment is lacking and the softening of state budget constraints is a serious problem (see, e.g., Rodden 2003).[3]

The present study is the first attempt to investigate the law for US and German states, with the exception of Abizadeh and Yousefi (1988), who study 10 US states during the period from 1950 to 1984 and provide support for the law. It should be noted that Abizadeh and Yousefi (1988) never concern themselves with whether the time series are stationary. The present study is also related to recent contributions by Narayan et al. (2008a, b, 2012), who conduct state-level analyses of Wagner's law for other countries. As emphasized in these studies, there are several reasons why one should examine Wagner's law at the state level.[4] Among others, important points in their discussion are summarized as follows.

First, given that Wagner did not take into account the influences of wars, a tacit assumption about the law is that the economy is under peacetime conditions. This is related to another leading theory, the so-called displacement effect (also known as the Peacock–Wiseman hypothesis), which can be relevant to the long-run behavior of government size under crises such as wars.[5] Although government expenditures are subject to political and military conflicts, their effects are lower at the state level than at the national level. From this perspective, the use of state-level data fits with Wagner's supposition. Second, when conducting cross-sectional or panel-data analysis, cultural and institutional differences across regions can be troublesome. Compared

[3] While the German federal government has prudent fiscal policies, state-level fiscal discipline is eroded. The "equivalence of living conditions" clause in the Basic Law compels the federal government to bail out a state that faces a debt crisis. Recent bailout episodes can be seen in the cases of Bremen and Saarland. See Rodden (2003) for more details on soft budget problems in German states.

[4] While almost all researchers investigate the law using national-level data, a recent direction in the literature on Wagner's law focuses on validity at the subnational or state level. Abizadeh and Yousefi (1988) produced the first paper to employ state-level data to test the law, using time-series data for 10 US states for the period 1950–1984, and their results support the law. More recently, applying a panel unit root, panel cointegration, and Granger causality analysis, Narayan et al. (2008a) examine the law on the basis of data from Chinese provinces and find mixed results. Narayan et al. (2008b) conduct time-series analysis for the Fiji islands and vindicate the law. Like Narayan et al. (2008a), utilizing panel-data techniques, Narayan et al. (2012) investigate the law for the 15 Indian states and provide strong support of the law.

[5] The displacement effect is initially found by Peacock and Wiseman (1961), who show that the sudden increase in government expenditure during World Wars I and II does not return to the pre-war levels in the UK. In other words, Peacock and Wiseman find stepwise increases in UK government size through World Wars I and II. Such a long-run growth in government size makes an analysis of Wagner's law difficult. Using historical data from Italy, Cavicchioli and Pistoresi (2016) find that military spending during wars results in nonlinearities between variables. See also Funashima (2017), who distinguishes between Wagner's law and the displacement effect.

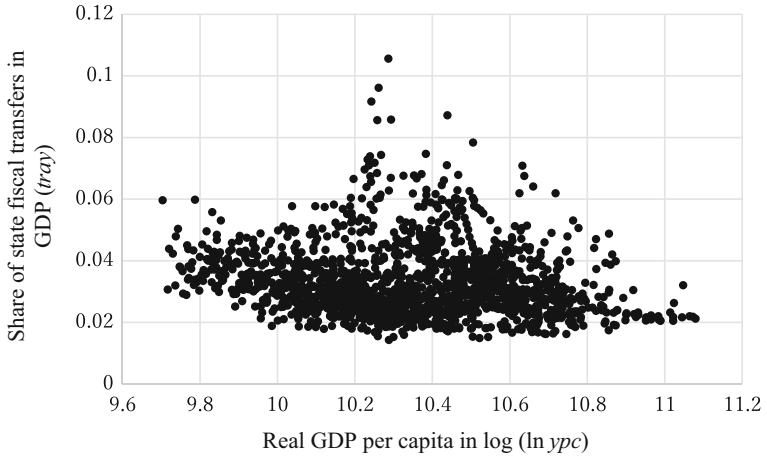

Fig. 1 Economic development and fiscal transfers in the US states. *Notes* The correlation coefficient is −0.1473. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1977 to 2010 for 47 states (excluding Alaska, Wyoming, and Hawaii)

with national- or federal-level data, the use of state-level data enables us to lessen the effects of such differences across regions. Third, a central (federal) government's expenditures are more likely to be influenced by international economic conditions than are a local government's expenditures.

Further, in contrast to previous studies, we attempt to undertake a more elaborate analysis by taking into account the effects of intergovernmental fiscal transfer payments on the validity of the law. As shown in Figs. 1 and 2, fiscal transfer is negatively correlated with economic development, indicating that fiscal transfer plays a role in horizontal equity across states, especially in the German federal system.[6] Fiscal transfer is, in fact, substantial in Germany, and the transfer dependence of the states is high. As such, Figs. 3 and 4 plot the relationship between fiscal transfer and state government sizes in the US and German states, respectively. The figures demonstrate that, in both countries, state government size is positively correlated with fiscal transfer, meaning that the latter may be a crucial factor in determining the former. In other words, it is highly probable that if fiscal transfers are ignored in empirical analyses of state-level Wagner's law, some omitted variable bias problems will occur and, consequently, misleading conclusions will be obtained.

Our empirical method is based on cointegration analysis, as in many recent predecessors, and incorporates extended panel-data techniques. Excluding Narayan et al. (2008a, 2012) and Lamartina and Zaghini (2011), this study is the first to apply the panel cointegration approach to test the law, while almost all previous studies use time-series data.

Our results uncover that, in both countries, fiscal transfers have non-negligible effects on state government size, and the introduction of fiscal transfers quantitatively

[6] It should be noted that, in early 1987, the German central government started to provide special supplementary transfers for the states of Bremen and Saarland, in order to handle their high debts.

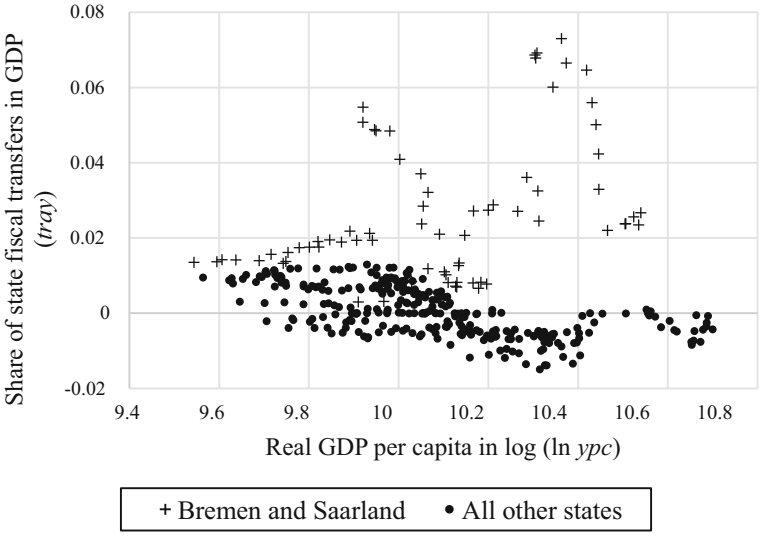

Fig. 2 Economic development and fiscal transfers in the German states. *Notes* The correlation coefficient is −0.0997 in the full sample and −0.6195 in the subsample excluding Bremen and Saarland. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1975 to 2010 for the 10 West German states (excluding Berlin)

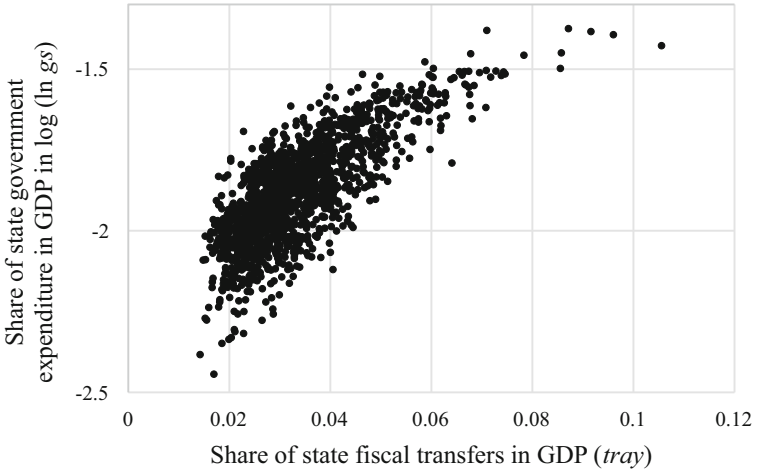

Fig. 3 Fiscal transfers and government size in the US states. *Notes* The correlation coefficient is 0.7689. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1977 to 2010 for 47 states (excluding Alaska, Wyoming, and Hawaii)

changes the validity of the law. This suggests that if fiscal transfers are omitted, then the outcomes are highly likely to be biased. However, the qualitative results are the same regardless of whether fiscal transfers are considered. That is, our results reveal that the law is validated in the US states, but, on the other hand, an "inverse" law is

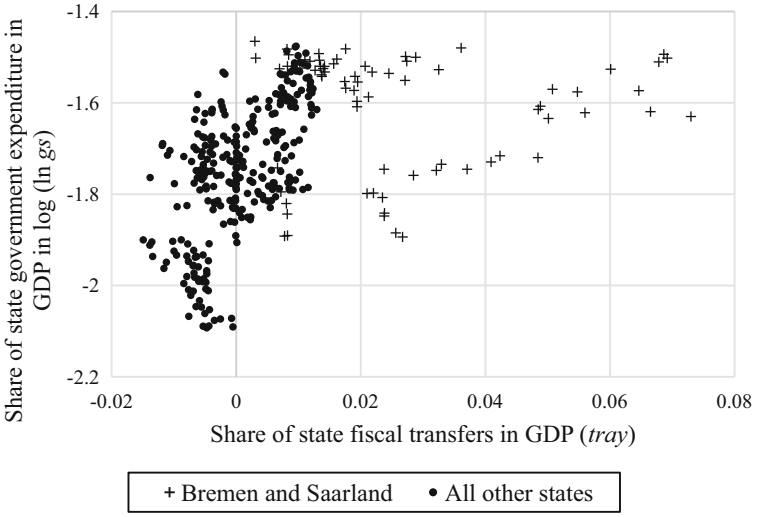

Fig. 4 Fiscal transfers and government size in the German states. *Notes* The correlation coefficient is 0.4737 in the full sample and 0.5963 in the subsample excluding Bremen and Saarland. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1975 to 2010 for the 10 West German states (excluding Berlin)

supported in the German states. Moreover, we find that the US validity of the law is weakened by some of the balanced budget requirements and that the German "inverse" law, a negative correlation between government size and output, is likely to be caused by the soft budget constraints, especially after the bailouts of Bremen and Saarland in 1992. These opposite outcomes between the US and German states imply that the characteristics of fiscal discipline are relevant to the validity of the law and that the public sector can grow through different channels. This implication offers new insights into the empirical literature, wherein heterogeneous consequences of the law are presented.

The rest of the paper is organized as follows: Section 2 briefly explains the federalism in the US and Germany including some stylized facts on intergovernmental transfers. Section 3 provides the present empirical framework and describes the data. Section 4 presents our empirical results including some robustness checks, and Sect. 5 concludes.

2 Federalism in the US and Germany

Before proceeding to our analysis, in what follows, we summarize the characteristics of federalism in the US and Germany following Bordo et al. (2013) and Potrafke and Reischmann (2015), who provide an extensive review of the characteristics.

2.1 The US

In the US, the fiscal system is highly decentralized and fiscal sovereignty is established in the states.

Subnational expenditures account for more than half of total expenditures and cover most major expenditures, except national defense, pensions, and health insurance for the elderly and disabled. Subnational revenue is almost half of total revenue, and the tax bases and rates are basically determined by the state government with discretionary powers.

The intergovernmental transfers are vertical from the federal to the state governments, mostly depending on expenditure by the states. The transfer dependence is not very large relative to that of German states. The federal grants to states are categorized into three types, that is, categorical, block, and general-purpose grants. The amount is calculated according to predetermined formulas or state-specific projects. The former refers to formula grants, which are calculated according to a formula based on population, per capita income, and other factors. The latter refers to project grants, which are allocated under a competitive system. With regard to the municipalities, they receive transfers from both the federal and state governments. The transfers from the federal government to municipalities are relatively small, as most of the transfers are from state governments.

In terms of fiscal discipline, although the borrowing autonomy of the states is not strongly limited, balanced budget rules are applied to them. In addition, the balanced budget requirements vary considerably across states (e.g., Mahdavi and Westerlund 2011). As Potrafke and Reischmann (2015) state, "some balanced budget requirements provide enough flexibility for states to carry over deficits if necessary" (p. 980). Further, the federal government forbids bailouts of states facing risk of default. Through the no-bailout lesson of 1840, the federal government signals a credible commitment to the no-bailout rule.

2.2 Germany

In Germany, when compared to the US, the shares of subnational expenditures and revenue of the total amounts are relatively low. The state expenditures are restrained by federal law. In addition, almost all tax bases and rates are determined by federal law, and tax rate autonomy is much less than in the US.

On the other hand, the intergovernmental transfer dependence of subnational government is much larger than in the US. Unlike in the US, the intergovernmental transfers are not only vertical from the federal to state governments, but also horizontal between state governments. The fiscal system is intended to guarantee a minimum level of tax revenues. The horizontal transfers mean that states with high per capita tax revenues pay transfers to states with low per capita tax revenues, harmonizing tax revenues across the states. In the vertical transfers, the federal government pays additional grants to states with low per capita tax revenues.[7] Incidentally, unlike the US case, in Germany, the municipalities only receive transfers from state governments.

[7] See Baretti et al. (2002) for details on the German fiscal equalization system. Note, however, that the current fiscal equalization system ("Länderfinanzausgleich") will disappear in 2019 and a new intergovernmental transfer system will be introduced in 2020. In other words, the new system differs substantially from the current system in that there are not direct horizontal transfers between states.

Importantly, the indiscipline of German local public finance is noted. Following the so-called golden rule until 2010, German states were officially only allowed to borrow for investment purposes, but in reality, they were able to simply circumvent the rule. Accordingly, it is recognized that the state governments had full autonomy in borrowing. In addition, the federal government officially follows the no-bailout rule, but the credibility of the commitment is lacking. In particular, the famous bailouts of the states of Bremen and Saarland worsen the indisciplined behavior of the states. As a result, the softening of state budget constraints is a serious problem.[8]

Incidentally, while our sample period does not cover it, the golden rule was recently replaced with the "debt brake," which was enshrined in the German Basic Law, and requires balanced budgets of central and state governments without incurring new debt.

3 Empirical framework and data

3.1 Panel cointegration analysis of Wagner's law for US and German states

As described in Ram (1987), Peacock and Scott (2000), and Durevall and Henrekson (2011), multiple variables have been hitherto supposed to test Wagner's law. Among others, the specification proposed by Musgrave (1969) is commonly used in empirical papers. In previous analyses, the share of government expenditure in gross domestic product (GDP) was used as a proxy for government size, and real income per capita was used as a proxy for economic development (see, e.g., Durevall and Henrekson 2011; Mann 1980).

Following the bulk of the recent empirical literature regarding Wagner's law, our analysis builds on cointegration analysis to investigate the long-run relationship between government size and economic development (e.g., Chang 2002; Durevall and Henrekson 2011; Islam 2001; Iyare and Lorde 2004; Kuckuck 2014; Lamartina and Zaghini 2011). As emphasized in Peacock and Scott (2000) and Lamartina and Zaghini (2011), a cointegrating relationship with a positive coefficient is consistent with the original view of Wagner, who did not consider a causality from economic development to government size, and vice versa.[9]

Since we use state-level panel data, our basic model can be written as:

$$\ln gs_{it} = \alpha_0 + \alpha_1 \ln ypc_{it} + \varepsilon_{it},$$

where gs is the share of state government expenditure in GDP (state government size), ypc is the real GDP per capita, and subscripts i and t are the cross section of states and time, respectively. Note that GDP is the total value of goods and services produced

[8] See Qian and Roland (1998) for the relationship between the decentralization of government and the soft budget constraint.

[9] Despite the original view, there is a strand of literature that investigates the causal relationship using Granger's causality tests (e.g., Chow et al. 2002; Iyare and Lorde 2004; Kuckuck 2014; Thornton 1999). Reflecting such a situation in the empirical literature, Lamartina and Zaghini (2011) utilize the Wald exogeneity test to examine the causality, in addition to a rigorous panel cointegration analysis.

within a state. If Wagner's law holds, then the coefficient α_1 should be significantly larger than zero.

As mentioned in our introduction, the present study considers the potential effects of fiscal transfer payments on the validity of state-level Wagner's law (i.e., the cointegrating relationship between *gs* and *ypc*).[10] To this end, by introducing the third variable, we also consider the modified specification:

$$\ln gs_{it} = \alpha_0 + \alpha_1 \ln ypc_{it} + \alpha_2 tray_{it} + \varepsilon_{it},$$

where *tray* denotes the share of state fiscal transfers in GDP (net transfers-to-GDP ratios). The coefficient α_2 should be significantly positive if fiscal transfers have positive effects on state government size.

As stated in the preceding section, although intergovernmental fiscal transfers are implemented to fund the budgets of state and local governments in the federal systems of both the US and Germany, a key difference in fiscal transfer systems exists between the two countries. In the US, the transfer payments are only vertical, and the federal government transfers to the states. However, in Germany, horizontal transfers between the states are performed in addition to vertical transfers. Hence, the US payments are basically positive, whereas the German payments can be negative in rich states.[11]

3.2 Datasets

All datasets are taken from Potrafke and Reischmann (2015) because their samples include fiscal transfers and collected as much data as is possible for the US and German states; thus, they are very useful to our study. For the US states, our annual data cover the period from 1977 to 2010 for 47 states, excluding Alaska, Wyoming, and Hawaii. For the German states, our annual data cover the period from 1975 to 2010 for the 10 West German states, excluding Berlin.[12]

Given Potrafke and Reischmann's (2015) datasets, *gs* is calculated as the share of nominal state government expenditure in nominal state domestic product, *ypc* is constructed by dividing real domestic product by population, and *tray* is calculated as the share of nominal net state fiscal transfers in nominal state domestic product. Table 1 shows the basic descriptive statistics for our datasets.

For *gs* and *tray*, we use the state-level data that include municipalities. Moreover, in order to focus on the state government behavior, as in Potrafke and Reischmann

[10] In the literature on Wagner's law, when conducting cointegration analysis, all studies suppose a bivariate system between government size and economic development. One notable exception is Chow et al. (2002), who emphasize the importance of controlling the effects of a third variable on the cointegrating relationship between government size and economic development.

[11] Net state fiscal transfers can be negative in Germany and ln *tray* is undefined. Baretti et al. (2002) focus on the German federal fiscal system and demonstrate that it is likely that the equalizing transfers reduce the tax revenue of the states.

[12] Following Potrafke and Reischmann (2015), these three US states are excluded because they are outliers. Likewise, Berlin is not included in our sample. Further, the East German states cannot be examined because of the lack of fiscal transfer data before 1995.

Table 1 Descriptive statistics

Variables	Mean	Median	Maximum	Minimum	Std. Dev.	Observations
The United States						
ln *ypc*	10.345	10.353	11.080	9.704	0.265	1598
Including municipalities						
ln *gs*	−1.880	−1.882	−1.375	−2.444	0.158	1598
tray	0.033	0.031	0.106	0.014	0.011	1598
Excluding municipalities						
ln *gs*	−2.308	−2.315	−1.554	−3.017	0.221	1598
tray	0.029	0.026	0.097	0.011	0.011	1598
Germany						
ln *ypc*	10.081	10.054	10.700	9.544	0.247	360
Including municipalities						
ln *gs*	−1.723	−1.724	−1.465	−2.093	0.149	360
tray	0.006	0.002	0.073	−0.015	0.015	360
Excluding municipalities						
ln *gs*	−2.055	−2.086	−1.465	−2.492	0.227	360
tray	0.006	0.002	0.073	−0.015	0.015	360

(2015), we also use data that exclude municipalities.[13] Figure 5 shows the scatter diagram of *ypc* and *gs* for the US states, including municipalities. A similar scatter diagram for the West German states is plotted in Figure 6. Although the individual state effects and other various factors are not controlled in these figures, one can see no evident correlation between state government size and real output per capita in the US states, while a negative correlation can be seen in the case of West German states. It should be noted that, in Germany, the relationship between *ypc* and *gs* becomes nonlinear in the full sample, largely because of the outliers of Bremen and Saarland.

4 Empirical results

4.1 Panel unit root tests

Prior to panel cointegration analyses, the first task is to check whether our variables contain a panel unit root. For the US states, our datasets include five variables: *ypc*, *gs* (including municipalities), *gs* (excluding municipalities), *tray* (including municipalities), and *tray* (excluding municipalities). For West German states, they include four variables: *ypc*, *gs* (including municipalities), *gs* (excluding municipalities), and *tray*. For these variables in levels and in first differences, we first perform two panel

[13] It should be noted that, at the German state level, the net transfers-to-GDP ratios (*tray*) are exactly the same, regardless of the inclusion of the municipalities, because the municipalities only receive transfers from state governments, as stated in the preceding section.

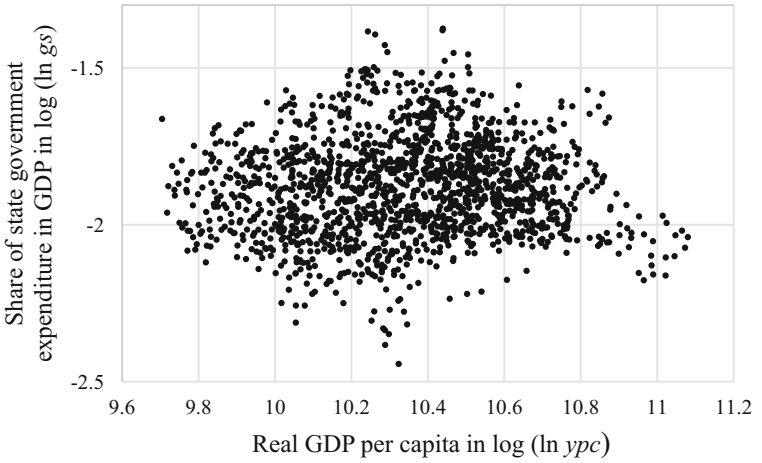

Fig. 5 Economic development and government size in the US states. *Notes* The correlation coefficient is 0.0562. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1977 to 2010 for 47 states (excluding Alaska, Wyoming, and Hawaii)

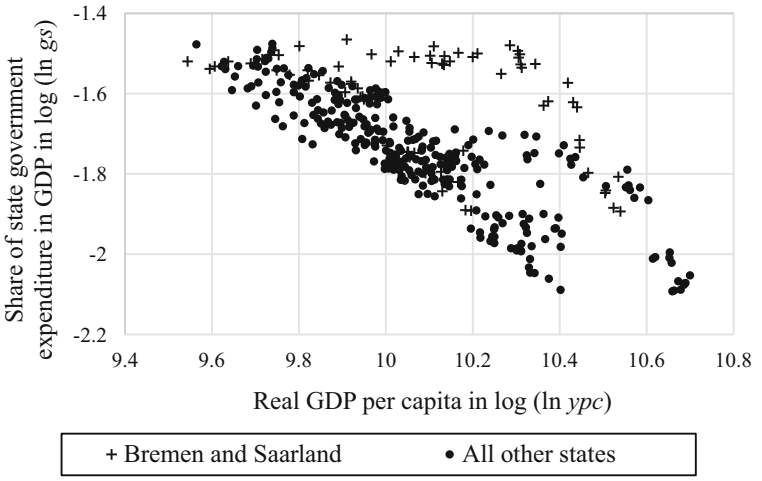

Fig. 6 Economic development and government size in the German states. *Notes* The correlation coefficient is -0.7183 in the full sample and -0.8572 in the subsample excluding Bremen and Saarland. GDP is the total value of goods and services produced within a state. Data are taken from Potrafke and Reischmann (2015), and the period is from 1975 to 2010 for the 10 West German states (excluding Berlin)

unit root tests: those from Levin et al. (2002) and Im et al. (2003). In the former test, homogeneity in the autoregressive coefficient across cross sections is assumed; the latter test allows for heterogeneity. The lag lengths are chosen based on the Schwarz information criterion (up to five lags). For all tests, we suppose two specifications: one includes constants and linear trends (detrended tests) and the other includes constants only (demeaned tests).

Table 2 Results of panel unit root tests

Variables	Levin et al. (2002)		Im et al. (2003)	
	No trend statistic	Trend statistic	No trend statistic	Trend statistic
The United States				
ln *ypc*	−5.687***	5.846	3.194	5.297
Δln *ypc*	−23.039***	−19.945***	−21.823***	−19.037***
Including municipalities				
ln *gs*	−0.034	−5.984***	1.804	−7.100***
Δln *gs*	−27.943***	−24.408***	−28.568***	−25.531***
tray	10.596	3.338	8.335	3.781
Δ*tray*	−14.363***	−11.338***	−18.433***	−18.708***
Excluding municipalities				
ln *gs*	−0.378	−6.841***	2.602	−6.866***
Δln *gs*	−32.728***	−28.230***	−32.444***	−28.370***
tray	12.605	4.028	11.848	0.917
Δ*tray*	−16.338***	−14.401***	−21.137***	−20.529***
Germany				
ln *ypc*	−7.536***	−3.124***	−3.362***	−1.666**
Δln *ypc*	−13.324***	−13.239***	−12.270***	−11.451***
Including municipalities				
ln *gs*	−0.823	−0.542	1.765	−1.118
Δln *gs*	−13.840***	−12.121***	−13.617***	−11.923***
tray	−0.279	−0.222	−0.507	−0.394
Δ*tray*	−19.919***	−17.537***	−17.951***	−16.164***
Excluding municipalities				
ln *gs*	−1.386*	−2.482***	0.941	−3.237***
Δln *gs*	−13.212***	−10.958***	−15.291***	−13.677***
tray	−0.279	−0.222	−0.507	−0.394
Δ*tray*	−19.919***	−17.537***	−17.951***	−16.164***

The lag lengths are chosen based on the Schwarz information criterion (up to five lags)
*, **, and *** represent the rejection of the null hypothesis at the 10, 5, and 1% significance levels, respectively

Table 2 reports the results. Overall, the null hypothesis of the unit roots in levels cannot be rejected at 1% significance levels in all cases of the tests of Levin et al. (2002) and Im et al. (2003), suggesting that each variable is panel non-stationary and has at least one panel unit root. On the other hand, one can confirm strong rejections for all of the variables in both tests when series are taken in first differences. We, therefore, determine that all of our underlying variables appear to be integrated of order 1 (I (1)).

4.2 Panel cointegration tests

In the next step, the panel cointegration tests proposed by Pedroni (1999) and Kao (1999) are conducted to examine whether there is a panel cointegrating relationship in the bivariate and trivariate systems: $(\ln gs, \ln ypc)$ and $(\ln gs, \ln ypc, tray)$, respectively. Both the tests of Pedroni (1999) and Kao (1999) are extensions of Engle–Granger's tests to treat panel data, based on an examination of the residuals. In Pedroni cointegration tests, we run four within-group and three between-group tests. The lag lengths are chosen based on the Schwarz information criterion (up to five lags). Regarding the deterministic components, we suppose the case of both trends and constants as well as of only constants in Pedroni cointegration tests, and only constants in Kao cointegration tests.

The Pedroni cointegration outcomes are reported in Table 3. In this table, we also present the German estimation results using subsamples excluding Bremen and Saarland, since observations for the two states can be considered large outliers causing nonlinearities between our variables, as shown in Figs. 2, 4, and 6. From Panel (A) of Table 3, in which the results of the bivariate case are shown, one can find that the null hypothesis of no cointegration is rejected in the US, on the whole. On the other hand, in the German case of the full sample, the results are somewhat unclear. When including municipalities, almost all results cannot strongly reject the null hypothesis in the full sample for German states. However, in the subsamples excluding Bremen and Saarland, the results of the ADF-statistic test with deterministic trends reject the null hypothesis at the 5% significance level, even when municipalities are included.

The results of the trivariate system are reported in Panel (B) of Table 3. Unlike the bivariate case, the presence of a panel cointegrating relationship in our three-variable system is strongly supported in not only the US, but also in the full sample and subsample for Germany, on the whole. In particular, almost all of the results of the ADF-statistic test reject the null hypothesis at the 1% significance level. This provides strong evidence for the presence of a panel cointegrating relationship in our three-variable system.

Table 4 displays the results of Kao cointegration tests. This indicates that the null is strongly rejected for all the cases.

To check the robustness of the above results on residual panel cointegration tests, we further implement an alternative test, the Johansen–Fisher cointegration test, which is an extension of Johansen's cointegration trace and maximum eigenvalue tests, to treat panel data. The Johansen–Fisher tests now include linear deterministic trends.

Table 5 presents the results of the Johansen–Fisher tests. Except for the German case including municipalities, both tests show that the null hypothesis of no cointegration is strongly rejected. On the whole, from the results of the Pedroni, Kao, and Johansen–Fisher cointegration tests, we can conclude that there is a panel long-run equilibrium relationship, at least in the trivariate system, for both the US and Germany.

4.3 Panel cointegrating vector estimations

Now we present the estimation results of the cointegrating vector. In order to estimate the vector, we utilize two methods: the fully modified ordinary least squares (FMOLS)

Table 3 Results of Pedroni cointegration tests

	Deterministic trend	Panel (within-dimension)				Group (between-dimension)		
		v-Stat	rho-Stat	PP-Stat	ADF-Stat	rho-Stat	PP-Stat	ADF-Stat
(A) *Two variables*								
The United States								
Including municipalities	Yes	17.729***	−1.927**	−4.285***	−5.838***	0.899	−3.548***	−5.475***
	No	7.584***	−3.882***	−3.485***	−6.650***	−1.378*	−2.759***	−6.953***
Excluding municipalities	Yes	11.022***	−1.928**	−4.835***	−5.951***	0.381	−5.681***	−6.067***
	No	7.095***	−4.809***	−4.579***	−6.345***	−2.479***	−4.398***	−6.984***
Germany								
Including municipalities	Yes	−0.589	0.471	−0.046	−1.161	1.389	0.624	−1.165
	No	1.481*	−1.046	−1.405*	−1.379*	0.242	−0.772	−0.909
Excluding municipalities	Yes	−0.471	0.428	−0.652	−1.883**	0.989	−0.986	−3.964***
	No	2.003**	−1.426*	−1.958**	−2.558***	−0.736	−2.434***	−4.126***
Germany (excluding Bremen and Saarland)								
Including municipalities	Yes	0.295	−0.254	−0.629	−2.217**	0.966	0.251	−1.716**
	No	1.732*	−1.321	−1.516*	−1.534*	0.006	−0.846	−1.023
Excluding municipalities	Yes	0.924	−0.946	−2.128***	−4.443***	0.450	−1.637*	−4.935***
	No	2.615**	−2.234***	−2.616***	−3.765***	−1.116	−2.793***	−4.729***
(B) *Three variables*								
The United States								
Including municipalities	Yes	9.734***	1.081	−1.351**	−2.658***	3.129	−1.595*	−4.202***
	No	4.533***	−0.624	−2.548***	−4.376***	0.884	−2.937***	−5.788***
Excluding municipalities	Yes	5.911***	−0.592	−3.593***	−4.996***	1.652	−4.287***	−6.248***
	No	5.747***	−2.127**	−4.640***	−5.730***	−0.632	−5.762***	−8.203***

Table 3 continued

	Deterministic trend	Panel (within-dimension)				Group (between-dimension)		
		v-Stat	rho-Stat	PP-Stat	ADF-Stat	rho-Stat	PP-Stat	ADF-Stat
Germany								
Including municipalities	Yes	−0.127	0.092	−0.932	−3.455***	1.181	−0.439	−3.150***
	No	1.186	−0.960	−1.698**	−2.627***	0.425	−0.917	−1.640*
Excluding municipalities	Yes	0.626	−0.563	−2.101**	−3.845***	0.872	−1.383*	−3.848***
	No	2.347***	−1.936**	−2.792***	−3.311***	−0.587	−2.445***	−3.240***
Germany (Excluding Bremen and Saarland)								
Including municipalities	Yes	−0.254	0.384	−0.635	−3.687***	1.285	−0.248	−3.253***
	No	0.954	−0.891	−1.488*	−2.531**	0.334	−0.769	−1.510*
Excluding municipalities	Yes	0.502	−0.300	−2.217***	−4.694***	0.930	−1.295*	−4.011***
	No	2.150	−2.349***	−3.097***	−3.757***	−0.801	−2.530***	−3.339***

The lag lengths are chosen based on the Schwarz information criterion (up to five lags)
*, **, and *** represent the rejection of the null hypothesis of no cointegration at the 10, 5, and 1% significance levels, respectively

Table 4 Results of Kao cointegration tests

	Statistic		Statistic
(A) *Two variables*		(B) *Three variables*	
The United States		The United States	
Including municipalities	−5.938***	Including municipalities	−9.379***
Excluding municipalities	−4.683***	Excluding municipalities	−8.856***
Germany		Germany	
Including municipalities	−4.083***	Including municipalities	−3.537***
Excluding municipalities	−3.050***	Excluding municipalities	−3.552***
Germany (excluding Bremen and Saarland)		Germany (Excluding Bremen and Saarland)	
Including municipalities	−4.181***	Including municipalities	−4.204***
Excluding municipalities	−3.167***	Excluding municipalities	−3.287***

The tests include constants only. The lag lengths are chosen based on the Schwarz information criterion (up to five lags)
*, **, and *** represent the rejection of the null hypothesis of no cointegration at the 10, 5, and 1% significance levels, respectively

and dynamic ordinary least squares (DOLS) methods. In what follows, while we report the results when the lag and lead lengths of the DOLS equation are assumed to be unity, the results when their lengths are two are almost the same.

The estimation results are reported in Table 6.[14] Panel (A) of Table 6 shows the results of the bivariate system, which indicate that the coefficients of ln *ypc*, estimated by FMOLS and DOLS, are positive and significant at the 1% significance level in the US case. In contrast, those of the German case are negative and significant at the 1% significance level, while their absolute values are vastly larger than those of the US case. The German full sample results are mirrored by our subsample analysis excluding Bremen and Saarland. These findings provide support for Wagner's law in the US states, but no support in the German states.[15]

Panel (B) of Table 6 shows the results of the trivariate system. In both countries, all of the estimated coefficients of *tray* are positive, and almost all of them are statistically significant.[16] This implies that fiscal transfers have positive effects on state government size. However, the qualitative validity of Wagner's law is robust in the sense that the signs of the coefficients of ln *ypc* do not depend on whether *tray* is taken into account. Hence, we can conclude that Wagner's law is strongly validated in the US states,

[14] In the DOLS results, heteroscedasticity- and autocorrelation-consistent (Newey–West) robust standard errors are used because of the autocorrelation of the residuals.

[15] See Koester and Priesmeier (2013) for the national-level relationship between Wagner's law and the sustainability of public finances in Germany.

[16] Not surprisingly, in Germany, the positive effects of fiscal transfers are statistically significant only when municipalities are included. This is because the municipalities only receive transfers from state governments and are exactly the same, regardless of the inclusion of the municipalities, as already mentioned. In other words, the municipalities receive a portion of *tray* from state governments and increase spending, but such a positive relationship between *tray* and ln *gs* is not reflected when municipalities are excluded.

Table 5 Results of Johansen–Fisher cointegration tests

	Lag	Trace statistic	Max-eigen statistic		Lag	Trace statistic	Max-eigen statistic
(A) *Two variables*				(B) *Three variables*			
The United States				The United States			
Including municipalities	1	211.6***	204.4***	Including municipalities	1	287.5***	221.0***
	2	185.4***	171.1***		2	317.7***	216.2***
	4	307.0***	247.4***		4	773.0***	539.5***
Excluding municipalities	1	191.1***	184.1***	Excluding municipalities	1	232.4***	194.2***
	2	203.8***	190.2***		2	300.0***	244.1***
	4	376.6***	323.3***		4	696.3***	537.4***
Germany				Germany			
Including municipalities	1	36.5**	27.9	Including municipalities	1	45.6***	43.4***
	2	32.6**	25.5		2	69.5***	63.8***
	4	28.0	20.6		4	98.2***	76.7***
Excluding municipalities	1	60.3***	52.9***	Excluding municipalities	1	39.5***	35.6**
	2	56.2***	48.1***		2	56.9***	48.8***
	4	51.1***	42.8***		4	113.8***	98.6***
Germany (excluding Bremen and Saarland)				Germany (excluding Bremen and Saarland)			
Including municipalities	1	31.2**	22.2	Including municipalities	1	40.7***	40.7***
	2	27.2**	19.5		2	61.8***	59.7***
	4	21.0	13.0		4	87.7***	69.4***
Excluding municipalities	1	54.9***	47.1***	Excluding municipalities	1	35.5***	32.6***
	2	50.3***	41.8***		2	50.6***	45.0***
	4	45.5***	36.6***		4	104.6***	91.9***

The tests include linear deterministic trends. Trace statistic and max-eigen statistic are Fisher statistics for trace and maximal eigenvalue tests of the null hypothesis of no cointegration, respectively. *, **, and *** represent the rejection of the null hypothesis of no cointegration at the 10%, 5%, and 1% significance levels, respectively

Table 6 Results of FMOLS and DOLS

	FMOLS		DOLS	
	ln *ypc*	*tray*	ln *ypc*	*tray*
(A) *Two variables*				
The United States				
Including municipalities	0.362***		0.339***	
	(0.014)		(0.013)	
Excluding municipalities	0.460***		0.437***	
	(0.015)		(0.014)	
Germany				
Including municipalities	−0.695***		−0.761***	
	(0.031)		(0.029)	
Excluding municipalities	−0.558***		−0.610***	
	(0.025)		(0.027)	
Germany (excluding Bremen and Saarland)				
Including municipalities	−0.701***		−0.774***	
	(0.033)		(0.028)	
Excluding municipalities	−0.548***		−0.603***	
	(0.023)		(0.022)	
(B) *Three variables*				
The United States				
Including municipalities	0.222***	5.991***	0.214***	4.131***
	(0.017)	(0.459)	(0.021)	(0.519)
Excluding municipalities	0.190***	10.455***	0.191***	9.150***
	(0.017)	(0.483)	(0.023)	(0.681)
Germany				
Including municipalities	−0.655***	7.374***	−0.711***	9.322***
	(0.031)	(1.746)	(0.037)	(2.035)
Excluding municipalities	−0.610***	1.185	−0.676***	1.710
	(0.023)	(1.194)	(0.025)	(1.205)
Germany (excluding Bremen and Saarland)				
Including municipalities	−0.636***	8.651***	−0.683***	11.003***
	(0.037)	(2.179)	(0.044)	(2.541)
Excluding municipalities	−0.595***	0.821	−0.655***	1.394
	(0.026)	(1.486)	(0.026)	(1.501)

The values in parentheses are the standard errors. In the DOLS results, heteroscedasticity- and autocorrelation-consistent (Newey–West) robust standard errors are used

*, **, and *** represent the rejection of the null hypothesis at the 10, 5, and 1% significance levels, respectively

whereas it does not hold in the German states. Such German results in the trivariate system are again mirrored by our subsample analysis excluding Bremen and Saarland.

4.4 Relationship between Wagner's law and fiscal discipline

In this subsection, we attempt to provide more detailed analyses of the role of fiscal (in)discipline in the validity of Wagner's law.

From the standpoint of fiscal discipline, in the US states, balanced budget rules at the state level can be relevant to the law. With the exception of Vermont, US state governments are obliged to follow various balanced budget requirements. It is possible that some balanced budget rules urge policymakers to coordinate expenditures with the degree of economic growth in the long run, although the rules are not stringent in the short run, as stated in Poterba (1996) and Potrafke and Reischmann (2015).

In order to capture the effects of fiscal stringency stemming from balanced budget rules, we identify the regional differences in balanced budget requirements across the US states. To do this, as in Mahdavi and Westerlund (2011, Table 1), we use the following five measures of the degree of fiscal stringency.[17]

The first measure is the Advisory Commission on Intergovernmental Relations' (ACIR's) fiscal stringency index (ACIR-FSI), which is based on the assessment of the ACIR (ACIR 1987, Table 3). ACIR-FSI represents the degree of stringency scale, and its values are between 0 and 10. The values are cumulated points, depending on whether the requirement is statutory or constitutional and on the nature of the requirement. A rough explanation of the data construct of ACIR-FSI is as follows.[18] In a state where the requirement is solely statutory, it receives 1 point. If the requirement is constitutional, it earns 2 points. If the governor only has to submit a balanced budget, it receives 1 point. If legislature only has to pass a balanced budget, it earns 2 points. If the carryover of a deficit may be done, but it must be corrected in next fiscal year, it earns 4 points. If a deficit cannot be carried over into next biennium (fiscal year), it earns 6 (8) points.

While the ACIR-FSI is an aggregate index of fiscal stringency, the remaining four measures (denoted by BBR2, BBR5, BBR7, and BBR9) are involved with more specific balanced budget requirements. All of them take 1 when a certain balanced budget requirement is in place, and 0 otherwise. A state where BBR2 is equal to 1 is forced to "balance the budget based on own-source revenues alone." In a state where BBR5 is equal to 1, "a limit is in place on the amount of debt." In a state where BBR7 is equal to 1, there exists "a control on supplementary appropriations" and "the opportunity to rebudget" is limited. Finally, in a state where BBR9 is equal to 1, "no deficit may be carried over to the next fiscal year or biennium."

Utilizing these five measures of balanced budget rules to split our observations, Table 7 explores the effects of balanced budget requirements on the validity of Wagner's law in US states. Regarding ACIR-FSI, we follow Mahdavi and Westerlund (2011) and divide the sample into two groups; in one group, ACIR-FSI $\geq$ 8, and in

[17] For details on the measures, see Mahdavi and Westerlund (2011).

[18] See ACIR (1987) for more details.

Table 7 Effects of balanced budget requirements in the US

	FMOLS		DOLS	
	ln *ypc*	*tray*	ln *ypc*	*tray*
Including municipalities				
ACIR-FSI ≥ 8	0.228***	5.274***	0.212***	3.285***
	(0.020)	(0.574)	(0.024)	(0.658)
ACIR-FSI ≤ 6	0.207***	7.866***	0.218***	6.342***
	(0.031)	(0.711)	(0.041)	(0.751)
BBR2 $= 1$	0.132***	4.750***	0.144**	1.312
	(0.047)	(1.294)	(0.062)	(1.671)
BBR2 $= 0$	0.247***	6.326***	0.233***	4.893***
	(0.018)	(0.467)	(0.020)	(0.481)
BBR5 $= 1$	0.247***	5.488***	0.236***	3.741***
	(0.022)	(0.712)	(0.023)	(0.771)
BBR5 $= 0$	0.203***	6.397***	0.196***	4.446***
	(0.025)	(0.599)	(0.033)	(0.702)
BBR7 $= 1$	0.147***	7.162***	0.116***	5.948***
	(0.031)	(0.774)	(0.040)	(0.922)
BBR7 $= 0$	0.265***	5.328***	0.269***	3.101***
	(0.020)	(0.571)	(0.023)	(0.623)
BBR9 $= 1$	0.187***	6.869***	0.218***	4.571***
	(0.055)	(1.303)	(0.075)	(1.712)
BBR9 $= 0$	0.230***	5.811***	0.213***	4.040***
	(0.017)	(0.485)	(0.020)	(0.518)
Excluding municipalities				
ACIR-FSI ≥ 8	0.196***	9.385***	0.194***	7.961***
	(0.022)	(0.598)	(0.029)	(0.848)
ACIR-FSI ≤ 6	0.173***	13.256***	0.184***	12.261***
	(0.026)	(0.773)	(0.035)	(1.067)
BBR2 $= 1$	0.063	10.935***	0.101	7.387***
	(0.047)	(1.107)	(0.067)	(1.714)
BBR2 $= 0$	0.224***	10.325***	0.216***	9.627***
	(0.018)	(0.535)	(0.023)	(0.731)
BBR5 $= 1$	0.227***	8.750***	0.227***	6.930***
	(0.021)	(0.664)	(0.022)	(0.829)
BBR5 $= 0$	0.160***	11.833***	0.162***	10.944***
	(0.026)	(0.689)	(0.037)	(1.033)
BBR7 $= 1$	0.109***	11.527***	0.096*	11.033***
	(0.035)	(0.918)	(0.050)	(1.396)
BBR7 $= 0$	0.235***	9.848***	0.245***	8.084***
	(0.019)	(0.549)	(0.022)	(0.716)

Table 7 continued

	FMOLS		DOLS	
	ln *ypc*	*tray*	ln *ypc*	*tray*
BBR9 = 1	0.093*	12.196***	0.147*	10.270***
	(0.055)	(1.017)	(0.081)	(1.706)
BBR9 = 0	0.210***	10.098***	0.200***	8.921***
	(0.018)	(0.543)	(0.022)	(0.742)

The values in parentheses are the standard errors. In the DOLS results, heteroscedasticity- and autocorrelation-consistent (Newey–West) robust standard errors are used
*, **, and *** represent the rejection of the null hypothesis at the 10, 5, and 1% significance levels, respectively

the other group, ACIR-FSI ≤ 6. Note that there are 34 and 13 states in the former and latter groups, respectively, and there are no states in which ACIR-FSI is equal to 7.

Focusing first on the results of ln *ypc* in the case of ACIR-FSI, regardless of whether municipalities are included, there are no systematic differences in outcomes. Further, it is suggested that BBR5 does not create a large difference in the coefficients of ln *ypc*. On the other hand, one can find that the balanced budget requirements of BBR2 and BBR7 yield smaller estimated values of α_1. Only in the case where municipalities are excluded, BBR9 substantially lowers the coefficients of ln *ypc*. These outcomes imply that some of the balanced budget requirements are effective for restraining the growth of government budget and size.

Turning now to the German case, in the preceding subsection, we have found a significantly negative correlation between ln *ypc* and *gs*, suggesting an "inverse" Wagner's law. In terms of the characteristics of fiscal discipline, one possible explanation for the "inverse" law would stem from the soft budget problems of the German states. If the credibility and commitment of the German federal government's no-bailout rule are lacking, then there would be an incentive for the state government to borrow excessively. In fact, as shown in Rodden (2005), the ossification of the bailout expectations of the German state governments is related to larger deficits, since they are less willing to cut politically painful expenditures in response to negative shocks, such that revenue is below expectation. Given that the softening of state budget constraints has a greater influence on poor or low-growth states than on rich or high-growth states, slack economic development in the long run would heighten government size.

To formally test this hypothesis, we estimate the cointegrating vector by dividing the German sample into two subsample periods: 1975–1992 and 1993–2010. The reason for adopting these subsamples is related to the famous episodes in 1992 when the bailout expectations of Bremen and Saarland were, ultimately, explicitly confirmed, although they started receiving the special funds in 1994 (see, e.g., Rodden 2003, 2005). It is quite probable that their bailouts have significantly softened German state budget constraints.

Table 8 reports the results of the estimation. Comparing the subsample outcomes, one can detect remarkable changes of the estimated coefficients of ln *ypc* between the two subsamples. Regardless of whether municipalities are included, the difference is

Table 8 Effects of bailouts of Bremen and Saarland in Germany

	FMOLS		DOLS	
	ln *ypc*	*tray*	ln *ypc*	*tray*
(A) *Two variables*				
Including municipalities				
1975–1992	−0.359***		−0.384***	
	(0.019)		(0.021)	
1993–2010	−1.223***		−1.373***	
	(0.070)		(0.083)	
Excluding municipalities				
1975–1992	−0.370***		−0.386***	
	(0.021)		(0.023)	
1993–2010	−0.885***		−1.010***	
	(0.052)		(0.058)	
Excluding Bremen and Saarland				
Including municipalities				
1975–1992	−0.406***		−0.437***	
	(0.021)		(0.024)	
1993–2010	−1.203***		−1.367***	
	(0.084)		(0.101)	
Excluding municipalities				
1975–1992	−0.445***		−0.472***	
	(0.024)		(0.027)	
1993–2010	−0.784***		−0.915***	
	(0.061)		(0.069)	
(B) *Three variables*				
Including municipalities				
1975–1992	−0.351***	2.873	−0.307***	−0.848
	(0.025)	(2.304)	(0.069)	(6.846)
1993–2010	−1.290***	−2.507	−1.241***	0.142
	(0.071)	(1.776)	(0.131)	(2.751)
Excluding municipalities				
1975–1992	−0.386***	7.985***	−0.340***	6.201
	(0.029)	(2.677)	(0.074)	(7.675)
1993–2010	−0.938***	−5.355***	−0.885***	−4.571**
	(0.054)	(1.297)	(0.113)	(2.297)
Excluding Bremen and Saarland				
Including municipalities				
1975–1992	−0.412***	4.747*	−0.397***	0.775
	(0.026)	(2.807)	(0.076)	(8.522)
1993–2010	−1.307***	−3.227	−1.296***	−0.318
	(0.084)	(2.215)	(0.158)	(3.431)

Table 8 continued

	FMOLS		DOLS	
	ln *ypc*	*tray*	ln *ypc*	*tray*
Excluding municipalities				
1975–1992	−0.468***	10.567***	−0.466***	9.357
	(0.028)	(3.229)	(0.082)	(9.554)
1993–2010	−0.872***	−6.831***	−0.844***	−6.197**
	(0.060)	(1.614)	(0.136)	(2.863)

The values in parentheses are the standard errors. In the DOLS results, heteroscedasticity- and autocorrelation-consistent (Newey–West) robust standard errors are used
*, **, and *** represent the rejection of the null hypothesis at the 10, 5, and 1% significance levels, respectively

really remarkable in both the bivariate and trivariate systems.[19] Comparing the former subsample of 1975–1992 to the full sample results, the estimated values of α_1 are shifted toward zero, meaning that real output per capita has a limited impact on government size. On the other hand, with the latter subsample of 1993–2010, substantially lower estimated values of α_1 are exhibited. Importantly, the full sample results are mirrored by our subsample analysis excluding Bremen and Saarland. This implies that stronger bailout expectations are driven by the bailouts of Bremen and Saarland in 1992, and consequently, the softening of state budget constraints is prevalent in many other states.

It is still important to examine several alternative subsample periods and check the robustness of the evidence. In this context, in addition to the rapidly increased revenues of Bremen and Saarland in 1994, Rodden (2005) points out considerable discontinuity in the mid-1990s: the reunified Berlin as well as the multiple new Eastern states joining the equalization system in 1995. Given these backgrounds, Rodden (2005) considers the two divided sample periods: the first period is pre-unification up to 1993 and the second period is post-unification from 1995 to 2003; the former sample includes 10 states of the West German federation (excluding Berlin) and the latter includes all 16 states. Although, following Rodden's (2005) approach, we use some alternative subsample periods such as 1975–1993 (or 1975–1994) and 1995–2010 to circumvent the discontinuity problem in the mid-1990s, many of the same results are obtained. Hence, our main findings are unaffected in these exercises.

In summary, while acknowledging that it is difficult to distinguish the effects of the bailouts clearly from those of the reunification, our results support the view that the "inverse" Wagner's law is more validated after the bailouts of Bremen and Saarland. Thus, we provide strong confirmation of the hypothesis that the German "inverse" law appears to be primarily explained by the softening of state budget constraints.

[19] When excluding municipalities, the coefficient of *tray* for 1993–2010 is significantly negative. This result arguably comes from the fiscal consolidation of the Maastricht Treaty as well as the bailout of Bremen and Saarland. In other words, for the purpose of fiscal consolidation, the state governments are forced to cut expenditures even when transfers increase.

5 Conclusion

In this paper, we examined Wagner's law for US and German states, which have different characteristics of fiscal discipline. This study sheds new light on the voluminous previous empirical works on Wagner's law in the following respects. First, to the best of our knowledge, the present study is the first attempt to investigate the law for US and German states, with the exception of Abizadeh and Yousefi (1988), who studied 10 US states. Second, our study differs from all of our predecessors that investigate the law at the state level in that fiscal transfers are taken into account. Third, although almost all previous studies use time-series data, this is the first to use the panel cointegration approach to test the law, except for Narayan et al. (2008a, 2012) and Lamartina and Zaghini (2011).

Our main findings are summarized as follows. First, we show that significant effects of fiscal transfers on state government size are shown in most samples of the two countries. While state-level data are useful to test the law, as mentioned earlier, this outcome provides an important suggestion that fiscal transfers should be taken into account when exploring Wagner's law at the state or subnational level. Second, we demonstrate that the characteristics of fiscal discipline are relevant to the validity of the law and that the qualitative results are the same regardless of whether fiscal transfers are considered. In other words, we provide ample evidence to support the law in US states and "inverse" law in German states. In the US results, it is suggested that some of the balanced budget requirements weaken the validity of the law.

Third, and most important, we uncover the German "inverse" law, especially after the bailouts of Bremen and Saarland. This finding points to a new channel of growth in government size: Soft budget constraints can cause significant negative correlation between government size and economic development. Such an intergovernmental moral hazard problem for fiscal management in German states would be relevant not only to other troubled subnational governments in many countries, such as Argentina and Brazil, but also to national-level issues, such as the discussion about the bailouts of Greece.

Although the relationship between Wagner's law and fiscal discipline appears to be overlooked in previous empirical works, our results suggest that it is essential to take into account the characteristics of fiscal discipline when evaluating the law. This suggestion offers new insights into the empirical literature wherein mixed results on the validity of the law are presented.

While beyond the scope of this paper, a question of interest is whether the recently introduced "debt brake" is effective in strengthening fiscal discipline in Germany. Taking into account our US results of the balanced budget requirements, a partial answer is that it is possible that the "debt brake" contributes to responsible fiscal behavior to some extent. If the data become utilizable, further analysis should be pursued in future research.

Acknowledgements We would especially like to thank two anonymous referees and Markus Reischmann for their valuable comments and suggestions. We would also like to thank Tomomi Miyazaki, Masaki Nakahigashi, Katsuyoshi Nakazawa, Kengo Nutahara, Ryo Okui, Eric Weese, Hideo Yunoue, and participants at the 72nd Annual Congress of the International Institute of Public Finance (August 2016) for their helpful comments. The usual disclaimers apply. Hiraga acknowledges the financial support from the

Japanese Ministry of Education, Culture, Sports, Science and Technology (Grant-in-Aid for Young Scientist (B), No.16K17135).

References

Abizadeh, S., & Yousefi, M. (1988). An empirical re-examination of Wagner's law. *Economics Letters*, *26*, 169–173.

Advisory Commission on Intergovernmental Relations, ACIR. (1987). Fiscal discipline in the federal system: National reform and the experience of the states. Washington, D.C.

Baretti, C., Huber, B., & Lichtblau, K. (2002). A tax on tax revenue: The incentive effects of equalizing transfers: Evidence from Germany. *International Tax and Public Finance*, *9*, 631–649.

Bordo, M. D., Jonung, L., & Markiewicz, A. (2013). A fiscal union for the Euro: Some lessons from history. *CESifo Economic Studies*, *59*, 449–488.

Cavicchioli, M., & Pistoresi, B. (2016). Testing threshold cointegration in Wagner's Law: The role of military spending. *Economic Modelling*, *59*, 23–31.

Chang, T. (2002). An econometric test of Wagner's law for six countries based on cointegration and error-correction modelling techniques. *Applied Economics*, *34*, 1157–1169.

Chow, Y. F., Cotsomitis, J. A., & Kwan, A. C. C. (2002). Multivariate cointegration and causality tests of Wagner's hypothesis: Evidence from the UK. *Applied Economics*, *34*, 1671–1677.

Durevall, D., & Henrekson, M. (2011). The futile quest for a grand explanation of long-run government expenditure. *Journal of Public Economics*, *95*, 708–722.

Funashima, Y. (2017). Wagner's law versus displacement effect. *Applied Economics*, *49*, 619–634.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*, 53–74.

Islam, A. M. (2001). Wagner's law revisited: Cointegration and exogeneity tests for the US. *Applied Economics Letters*, *8*, 509–515.

Iyare, S. O., & Lorde, T. (2004). Co-integration, causality and Wagner's law: Tests for selected Caribbean countries. *Applied Economics Letters*, *11*, 815–825.

Kao, C. D. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, *90*, 1–44.

Koester, G. B., & Priesmeier, C. (2013). Does Wagner's law ruin the sustainability of German public finances? *FinanzArchiv: Public Finance Analysis*, *69*, 256–288.

Kolluri, B. R., Panik, M. J., & Wahab, M. S. (2000). Government expenditure and economic growth: Evidence from G7 countries. *Applied Economics*, *32*, 1059–1068.

Kuckuck, J. (2014). Testing Wagner's law at different stages of economic development. *FinanzArchiv: Public Finance Analysis*, *70*, 128–168.

Lamartina, S., & Zaghini, A. (2011). Increasing public expenditure: Wagner's law in OECD countries. *German Economic Review*, *12*, 149–164.

Levin, A., Lin, C. F., & Chu, C. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, *108*, 1–24.

Mahdavi, S., & Westerlund, J. (2011). Fiscal stringency and fiscal sustainability: Panel evidence from the American state and local governments. *Journal of Policy Modelling*, *33*, 953–969.

Mann, A. J. (1980). Wagner's law: An econometric test for Mexico, 1925–1976. *National Tax Journal*, *33*, 189–201.

Musgrave, R. A. (1969). *Fiscal systems*. New Haven, CT, and London: Yale University Press.

Narayan, P. K., Nielsen, I., & Smyth, R. (2008a). Panel data, cointegration, causality and Wagner's law: Empirical evidence from Chinese provinces. *China Economic Review*, *19*, 297–307.

Narayan, P. K., Prasad, A., & Singh, B. (2008b). A test of the Wagner's hypothesis for the Fiji islands. *Applied Economics*, *40*, 2793–2801.

Narayan, S., Rath, B. N., & Narayan, P. K. (2012). Evidence of Wagner's law from Indian states. *Economic Modelling*, *29*, 1548–1557.

Oates, W. (2005). Toward a second-generation theory of fiscal federalism. *International Tax and Public Finance*, *12*, 349–373.

Peacock, A. T., & Scott, A. (2000). The curious attraction of Wagner's law. *Public Choice*, *102*, 1–17.

Peacock, A. T., & Wiseman, J. (1961). *The growth of public expenditure in the United Kingdom*. Princeton, NJ: Princeton University Press.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, *61*, 653–670.

Poterba, J. M. (1996). Budget institutions and fiscal policy in the U.S. states. *American Economic Review Papers and Proceedings*, *86*, 395–400.

Potrafke, N., & Reischmann, M. (2015). Fiscal transfers and fiscal sustainability. *Journal of Money, Credit and Banking*, *47*, 975–1005.

Qian, Y., & Roland, G. (1998). Federalism and the soft budget constraint. *American Economic Review*, *88*, 1143–1162.

Ram, R. (1987). Wagner's hypothesis in time-series and cross-section perspectives: Evidence from "real" data for 115 countries. *Review of Economics and Statistics*, *69*, 194–204.

Rodden, J. (2003). Soft budget constraints and German federalism. In J. A. Rodden, G. S. Eskeland, & J. I. Litvack (Eds.), *Fiscal decentralization and the challenge of hard budget constraints* (pp. 161–186). Cambridge, MA: MIT Press.

Rodden, J. (2005). And the last shall be first: Federalism and soft budget constraints in Germany. Mimeo (Massachusetts Institute of Technology).

Rowley, C. K., & Tollison, R. D. (1994). Peacock and Wiseman on the growth of public expenditure. *Public Choice*, *78*, 125–128.

Shelton, C. A. (2007). The size and composition of government expenditure. *Journal of Public Economics*, *91*, 2230–2260.

Thornton, J. (1999). Cointegration, causality and Wagner's law in 19th century Europe. *Applied Economics Letters*, *6*, 413–416.

Weingast, B. R. (2009). Second generation fiscal federalism: The implications of fiscal incentives. *Journal of Urban Economics*, *65*, 279–293.