Tax revenue and financial development

Scott Gilbert
Department of Economics, SIUC
Carbondale, IL
gilberts@siu.edu

Bojan Ilievski
Department of Economics, SIUC
Carbondale, IL
bojan@siu.edu

January 20, 2012

Abstract

This paper considers the role of the financial sector, specifically banks, as a mechanism for tax enforcement and collection. We explore the hypothesis that the public sector, measured by the tax-to-GDP ratio, co-emerges with the banking sector, measured by deposits-to-GDP, during the course of economic development. We examine this hypothesis theoretically, tracing out various paths along which the banking and public sectors might co-emerge due to changes in tastes or technology. We then test it empirically, using panel data on 116 countries over the period 1990-2008. Evidence supports the hypothesis.

Keywords: tax, public good, bank deposits, Nash equilibrium, panel data
JEL Codes: H2 (taxation), G2 (financial institutions)
1 Introduction

Why is it that tax revenue as a share of GDP is significantly lower in developing economies than in developed economies? A simple answer is taste – in developing countries the need or desire for public works may be less than in developed countries. In support of this demand-side view, Tanzi and Zee (2000) argue that there is no need to focus on raising the level of tax revenue because “economic development would more often than not generate additional needs for tax revenue to finance the rise in public expenditure while at the same time increasing the countries’ ability to raise revenue to meet those needs.”

On the supply side, lower tax-to-GDP ratios in developing economies may reflect difficulty in extracting tax due to weak methods of tax enforcement and collection. Specifically, the existence of an informal sector or “shadow economy” can thwart efforts at taxation. An extensive literature supports this idea, including Piggott and Whalley (2001), Stiglitz (2002), and Emran and Stiglitz (2005).\textsuperscript{1} Recently Gordon and Li (2009) have argued that, as the financial sector improves its effectiveness, firms have increased incentive to make use of financial intermediaries. This in turn makes the formal sector bigger and raises the potential for the government to increase tax revenue.

With a large informal sector in developing countries, both demand-side and supply-side shifts (tastes and technology) may bring about a shrinkage in this sector and an expansion of taxable economic activity, thereby raising tax relative to GDP. The present work considers this proposition in a framework where taxable activity occurs via financial intermediaries (“banks”). With increasing reliance on banks along the path of economic development, tax collection becomes easier, and the tax-to-GDP ratio goes up. This suggests that the size of both the banking sector and public sector, relative to GDP, increase together during

\textsuperscript{1}Auriol and Warlters (2005) argue, on the other hand, that tax-seeking governments can actually seek to increase the size informal sector when firms face high fixed costs. In their theory, government-imposed costs of entry to the formal sector create market power and rents for firms in the formal sector.
economic development. We call this the co-emergence hypothesis.

To give more economic substance to the hypothesis of bank-tax co-emergence, we first examine it theoretically via models with two goods – public and private – and two forms of payment – checks and cash. In equilibrium we show that the co-emergence of bank and tax can arise via several channels: an increased preference for the public good, decreased necessity of cash, increased benefit of checks, and increased efficiency of public good provision. These results hold when the government’s marginal tax rate is held fixed, and also when it is endogenous.

We next examine the bank-tax co-emergence hypothesis empirically, using a panel of 116 countries over the period 1990-2008. We test whether tax is positively linked to bank deposits (each expressed as a share of GDP). The idea here is not that an increase in bank deposits necessarily causes an increase in the tax-to-GDP ratio per se, but that underlying factors – including those discussed in our theoretical framework – cause them to move together. If we could observe these underlying factors (tastes, technology, etc.) then we could build a causal empirical model of public finance, but this is currently infeasible. We can however estimate the level of tax-to-GDP conditional on bank deposits and some controls, including GDP and international trade, that have known historical relationships with tax-to-GDP.

We estimate models of tax-to-GDP (log-)levels, and growth rates, via panel econometric models in various ways – all instances of the generalized method of moments. We consistently find a significant positive link between the banking and public sectors, hence the data support the co-emergence hypothesis. To the extent that the government plays a role in maintaining the health of banking institutions, our research highlights the potential importance of that role for a different objective: raising tax and hence funding the general mission of the public sector.

The remainder of the paper is organized as follows. Section 2 briefly discusses the historical relationship between tax levels and bank deposits, Section 3 considers tax-bank co-
emergence in theory, Section 4 describes the data and reports empirical results, and Section 5 concludes. Proofs of theoretical economics results are contained in Appendix A1, and some relevant econometric details are derived in Appendix A2. All data and computer code used in the paper are available online at the corresponding author’s website.

2 Tax and Banks

In poorer countries tax revenues tend to be smaller as a share of GDP, and the same is true of bank deposits. For the years 1990-2008, Table 1 shows that in low-income and middle-income countries the tax-to-GDP ratio is about half the deposits-to-GDP ratio, with each ratio increasing in GDP. For high income countries, tax-to-GDP is about a third of deposits-to-GDP. ²

The great disparity in the size of the banking sector, across stages of economic development, underscores the problem in poor countries of monitoring and taxing economic activity. We focus on banks as a proxy for the general financial sector and for the “formal” economy. As the public sector expands over the course of development, tax revenues (relative to GDP) typically rise, but taxes change in other ways too. Specifically, the share of tax revenues derived from income tax rises, while that based on corporate tax and border taxes falls.³ To simplify matters we do not disaggregate tax into its sources, hence we ignore any changes in the tax structure along the development path.

²Results are based on the authors’ calculations. We divide economies according to 2010 GDP per capita, calculated using the World Bank Atlas method. For each group we compute averages using GDP as weight, for the panel data discussed in Section 4, with 116 countries observed during the years 1990-2008.
³See Table 1 of Gordon and Li (2009).
3  Theory

Suppose that households spend income via two means: cash and checks. Cash spending is unobserved by the government and hence not taxed, while check spending is taxed at a rate \( \tau_b \) for each dollar spent. With \( n \) households, let \( S_{ib} \) be household \( i \)'s spending via banks ("checks"), in which case total tax raised \( T \) is:

\[
T = \tau_b \sum_{i=1}^{n} S_{ib} \tag{3.1}
\]

Let \( T \) be total tax divided by total income, and let \( B \) total check spending divided by income. We can then recast the tax equation (3.1) as:

\[
T = \tau_b B \tag{3.2}
\]

If we suppose that spending via banks equals some fixed multiple of bank deposits then the relationship (3.2) is a simple interpretation of what is going on in the historical period described by Table 1. As the banking sector gets bigger so does taxable income and the amount of tax. This interpretation is valid when the marginal tax rate \( \tau_b \) is held fixed at some (exogenous) value, and would still be valid if \( \tau_b \) rose with economic development, as then \( T \) would swell due to increases in \( B \) and also \( \tau_b \).

Missing from (3.2) is any sense of how or why banking \( B \) and tax \( T \) might vary from one country to the next, or one year to the next. In the following sub-sections we attempt to account for a variety of circumstances – all likened to changes in “tastes” or “technology” – which might cause such variation.

3.1 A Simple Model

Consider a simple economy with \( n \) households and two goods: good \#1 a private good and good \#2 a public good – non-rival and non-excludable. Household \( i \)'s consumption levels
are $C_{i1}$ and $C_{i2}$ for the two goods, and since the second good is a public good the value of $C_{i2}$ is the same for each $i$, call it $C_2$. Households derive utility $U$ from consumption, of Cobb-Douglas form:

$$U_i = C_{i1}^{1-\alpha}C_2^\alpha$$ (3.3)

with parameter $\alpha$ in the range $(0,1)$. To acquire consumption goods, each household $i$ has income $Y_i$ which it spends on the private good, and pays tax with which the government commissions the provision of the public good.

Suppose that households spend income via two means: cash and checks. Cash spending incurs a loss which we will call “shrinkage” but proxies for a number of factors, including the opportunity costs of lost interest and efficiency associated with deposit-based spending. Cash spending is unobserved by the government and hence not taxed. Check spending does not incur shrinkage, but does incur tax. Denoting $S_a$ and $S_b$ the spending levels via cash and check, respectively, total spending equals income:

$$S_{ia} + S_{ib} = Y$$ (3.4)

and the household’s consumption of the private and public goods are:

$$C_{i1} = (1 - \tau_a)S_{ia} + (1 - \tau_b)S_{ib}$$ (3.5)

$$C_2 = \tau_b \sum_{j=1}^{n} S_{jb}$$ (3.6)

with $\tau_a$ the per-unit shrinkage loss associated with cash spending, and $\tau_b$ the government’s sales tax on the private good, each taking values in $(0,1)$. Households observe the values of $\tau_a$ and $\tau_b$ and select their spending levels $S_{ia}$ and $S_{ib}$ so as to maximize utility $U_i$ of
consumption, taking as given the choices of the remaining households.

To study outcomes economy-wide, let \( S_a \) and \( S_b \) be per-capita spending via cash and checks: \( S_a = n^{-1} \sum_i S_{ia} \) and \( S_b = n^{-1} \sum_i S_{ib} \). In equilibrium, check and cash spending behavior is as follows:

**Proposition 1.** For the economy defined by (3.3)-(3.6), in Nash equilibrium with spending via both cash and checks, the amount of check spending per capita \( S_b \) is:

\[
S_b = \frac{1 - \tau_a}{(\tau_b - \tau_a)(1 + n \frac{1-\alpha}{\alpha})} Y \tag{3.7}
\]

while cash spending is \( S_a = Y - S_b \). Here check spending becomes larger and cash spending becomes smaller in each of the following scenarios: (i) preference \( \alpha \) for the public good is higher, (ii) cash shrinkage \( \tau_a \) is higher, (iii) tax rate \( \tau_b \) is lower, (iv) population \( n \) is smaller.

Proposition 1 gives a simple economic context in which to interpret the bank-tax relationship (3.2). In any of the scenarios (i)-(iv), the deposits-to-GDP ratio \( B = \frac{S_b}{Y} \) rises, while at the same time the tax-to-GDP ratio \( T = \tau_b \frac{S_b}{Y} \) also rises. The scenarios include a shift in taste (\( \alpha \)) for the public good, and also shifts in costs – including the “shrinkage” cost of cash use and the tax on check use.

Proposition 1 takes as given that households spend via both checks and cash, and for this the ratios \( \frac{S_{ib}}{Y} \) must all lie in the range (0,1). In equilibrium, each of these ratios equals \( \frac{S_b}{Y} \), and from equation 3.7 this latter value lies in (0,1) if and only if:

\[
\tau_b > \tau_a + \frac{1 - \tau_a}{1 + n \frac{1-\alpha}{\alpha}} \tag{3.8}
\]

Hence households use both cash and checks if the tax rate \( \tau_b \) is not too low. Denoting by \( \tau_b^* \) the right-hand side of (3.8), if instead \( \tau_b \leq \tau_b^* \) then there is increased incentive to use checks, and a Nash equilibrium exists in which all households use only checks, no cash.
Another reason that both the banking and public sector may remain small is that some purchases are necessarily cash-only, yielding a cash constraint:

\[
\frac{S_{ia}}{Y_i} \geq \gamma
\]

for some value \( \gamma \) in \((0, 1)\) and all households \( i \). If the economy without a cash constraint is in equilibrium with spending via check and cash then, upon introducing a cash constraint, the constraint is binding on a per-capita basis if and only if:\(^4\)

\[
\gamma > 1 - \frac{1 - \tau_a}{(\tau_b - \tau_a)(1 + n^{1-\alpha})}
\]

Here a Nash equilibrium exists in which the cash constraint is binding for each household: \( \frac{S_{ia}}{Y} = \gamma \) for each household \( i \). If the fraction \( \gamma \) of cash-required spending drops as economies develop then this increases tax \( T \) and banks \( B \), provided that the constraint is binding. The diminishing cash constraint is therefore another channel through which banks and tax can co-emerge.

### 3.2 Endogenous Tax Rate

Suppose that the government, knowing the households’ intentions, selects the tax rate \( \tau_b \) in order to promote social welfare \( W \) defined as an increasing function \( W(U_1, ..., U_n) \) of household utilities.

In equilibrium, if households spend via both cash and checks then per capita spending levels are specified by Proposition 1, and since agents are identical their spending levels \( S_i \) each equal the per capita levels \( S \). Using equation 3.7 for check spending \( S_{ib} \), and the value \( Y - S_{ib} \) for cash spending, we plug spending levels into consumption equations 3.5-3.6 to get

\(^4\)This follows from equation 3.7 and the fact that \( S_{ia} + S_{ib} = Y \).
equilibrium consumption for the typical household:

\[ C_{i1} = (1 - \tau_a) \left(1 - \frac{1}{1 + n \frac{1-\alpha}{\alpha}}\right) Y \quad (3.11) \]

\[ C_2 = \frac{n \tau_b (1 - \tau_a)}{\tau_b - \tau_a} \frac{\tau_a}{(1 + n \frac{1-\alpha}{\alpha})} Y \quad (3.12) \]

Here private good consumption \( C_{i1} \) is unaffected by the tax rate \( \tau_b \). For public good consumption \( C_2 \) recall that households spend via cash and checks only if \( \tau_b > \tau_b^* \), in which case \( \tau_b > \tau_a \) because \( \tau_b^* > \tau_a \). The sign of the tax effect \( \frac{\partial C_2}{\partial \tau_b} \) is the same as that of \( \frac{\partial}{\partial \tau_b} \frac{\tau_a}{\tau_b - \tau_a} \), and the latter derivative equals \(-\frac{\tau_a}{(\tau_b - \tau_a)^2}\) which is negative. Hence public good consumption \( C_2 \) decreases with higher \( \tau_b \), while \( C_{i1} \) is constant in \( \tau_b \). Utility \( U_i \) and welfare \( W \) are therefore decreasing in \( \tau_b \) over the range \((\tau_b^*, 1)\).

If instead households spend only via checks then:

\[ C_{i1} = (1 - \tau_b)Y \quad (3.13) \]
\[ C_2 = n \tau_b Y \quad (3.14) \]
\[ U_i = Y (1 - \tau_b)^{1-\alpha} (n \tau_b)^\alpha \quad (3.15) \]

Here utility \( U_i \) is increasing in \( \tau_b \) when \( \tau_b \) is sufficiently small, reaching a maximum over \( \tau_b \) in \([0, \tau_b^*]\) either at \( \tau_b^* \) or at \( \alpha \), whichever is smaller.\(^5\) Hence social welfare \( W \) increases in \( \tau_b \) over the range \([0, \tau_b^*]\) when \( \tau_b^* \leq \alpha \), but peaks at \( \alpha \) when \( \alpha < \tau_b^* \).

Piecing together the above-described relationships between welfare \( W \) and tax rate \( \tau_b \) yields the following:

\(^5\)With \( U_i = Y (1 - \tau_b)^{1-\alpha} (n \tau_b)^\alpha \), its unconstrained maximum with respect to \( \tau_b \) is achieved at \( \tau_b = \alpha \).
Proposition 2. For the economy defined by (3.3)-(3.6), with Nash equilibrium and spending via cash and check, the social planner can always increase welfare $W$ by lowering the tax rate $\tau_b$ toward $\tau_b^*$. If instead households spend only via checks then the social planner maximizes welfare $W$ by setting the tax rate $\tau_b$ equal to either $\tau_b^*$ or $\alpha$, whichever is smaller.

According to Proposition 2, the social planner should choose a tax rate close to $\tau_b^*$ or $\alpha$. To interpret this result suppose that, along the course of economic development, one or more of the following situations takes place: preference $\alpha$ for the public good increases, shrinkage loss $\tau_a$ due to check spending rises, or population $n$ falls. In each case, the target(s) for the tax rate $\tau_b$ are rising.

With binding cash constraint (3.9), the social planner can not achieve check-only spending. If all households are cash constrained, each spending $S_{ia} = \gamma Y$ via cash and $S_{ib} = (1 - \gamma)Y$ via check, then consumption and utility are:

$$C_{i1} = ((1 - \tau_a)\gamma + (1 - \tau_b)(1 - \gamma))Y$$

$$C_2 = n\tau_b(1 - \gamma)Y$$

$$U_i = ((1 - \tau_a)\gamma + (1 - \tau_b)(1 - \gamma))^{1-\alpha}(n\tau_b(1 - \gamma))^{\alpha}Y$$

The tax rate that maximizes utility specified by (3.18) is:

$$\tau_b = \frac{\alpha}{1 - \gamma} \frac{1 - \gamma \tau_a}{1 - \gamma}$$

but with a binding cash constraint this tax rate may not be feasible as $\tau_b$ must satisfy (3.10). Rewriting (3.10), the relevant bound on $\tau_b$ is:

$$\tau_b < \tau_a + \frac{1 - \tau_a}{(1 - \gamma)(1 + n1 - \alpha)}$$
Hence, with a binding cash constraint the (constrained) optimal tax rate is the smaller of the values appearing on the right-hand sides of equations 3.19 and 3.20.

3.3 Consumption Elasticity

Our simple model in Section 3.1 assumed unit elasticity in the consumption of private and public goods. To consider the bank-tax relationship in somewhat more general terms let utility exhibit constant elasticity of substitution (CES):

\[ U_i = ((1 - \alpha)C_{i1}^\theta + \alpha C_{i2}^\theta)^{\frac{1}{\theta}} \]  

(3.21)

for some constant \( \theta < 1 \), in which case elasticity is \( \eta = \frac{1}{1-\theta} \), reducing to Cobb-Douglas utility and unit elasticity when \( \theta \to 0 \).

**Proposition 3.** For the economy defined via budget constraints (3.5) - (3.6) and CES consumption preferences (3.21), if all households spend via cash and check then equilibrium check spending per capita is:

\[ S_b = \frac{1 - \tau_a}{\tau_b - \tau_a + n\tau_b \left( \frac{1 - \alpha}{\alpha} \frac{\eta - \tau_a}{\tau_b} \right)} \eta Y \]  

(3.22)

and behaves as described in scenarios (i)-(iv) of Proposition 1, regardless of the consumption elasticity \( \eta \) between private and public goods.

By introducing elasticity \( \eta \) as a separate parameter, it raises the question of what would happen to spending patterns (cash versus check) if elasticity changed with economic development. From (3.22), the answer depends on the value of \( \frac{1 - \alpha}{\alpha} \frac{\eta - \tau_a}{\tau_b} \): if this value is less than 1 then an increase in elasticity also increases check spending, while if greater than 1 the opposite holds.
3.4 Technology

In Section 3.1 we also assumed, albeit implicitly, that a given dollar could purchase/generate equal amounts of public and private goods. This is consistent with the idea that households are endowed with some resource ("money"), and that they use it with constant returns to scale, such that returns are the same for private and public good creation. More generally let \( \lambda_1 \) and \( \lambda_2 \) be the (constant) returns to investment in private and public goods, respectively, so that:

\[
C_{i1} = \lambda_1((1 - \tau_a)S_{ia} + (1 - \tau_b)S_{ib})
\]

\[
C_2 = \lambda_2\tau_b \sum_i S_{ib}
\]

**Proposition 4.** For the economy defined via budget constraints (3.5) - (3.6) and CES consumption preferences (3.21) and productivity (3.23)-(3.24), if all households spend via cash and check then equilibrium check spending per capita is:

\[
S_b = \frac{1 - \tau_a}{\tau_b - \tau_a + n\tau_b \left( \frac{\lambda_1}{\lambda_2} \right)^{\eta-1} \left( \frac{1 - \alpha \tau_b - \tau_a}{\alpha \tau_b} \right)^{\eta} Y}
\]

This behaves as described in scenarios (i)-(iv) of Proposition 1, regardless of the productivity values \( \lambda_1 \) and \( \lambda_2 \).

If we imagine further the spending behavior that results from a change in productivity along the course of economic development, an increase in the relative productivity \( \frac{\lambda_2}{\lambda_1} \) of public good production increases bank spending \( S_b \) if elasticity \( \eta > 1 \), and decreases \( S_b \) if \( \eta < 1 \).
3.5 Income Inequality

We can also incorporate income inequality into the framework of Section 3.1. The result is as follows:

**Proposition 5.** For the economy defined by (3.3)-(3.6) with household incomes $Y_1, \ldots, Y_n$, if all households spend via both cash and check then the typical household’s equilibrium check spending is:

$$S_{ib} = cY_i + d \sum_{j \neq i} Y_j$$

and cash spending is $Y_i - S_{ib}$, with constants $c$ and $d$ defined as:

$$c = \frac{1 - \tau_a}{\tau_b - \tau_a} \left( 1 - \frac{1}{n + \frac{\alpha}{1 - \alpha}} \right)$$

$$d = -\frac{1 - \tau_a}{\tau_b - \tau_a} \left( \frac{1}{n + \frac{\alpha}{1 - \alpha}} \right)$$

In aggregate, the relationship between per capita check spending $S_b$ and per capita income $Y$ is invariant to the distribution of income across households.

We can also reconsider the idea of an endogenous tax rate $\tau_b$, here in the context of income inequality, with the following result:

**Proposition 6.** Under the assumptions of Proposition 5, consumption levels are the same for each household, given by:

$$C_{i1} = \frac{1 - \tau_a}{n + \frac{\alpha}{1 - \alpha}} \sum_j Y_j$$

$$C_2 = \frac{\tau_b(1 - \tau_a)}{(\tau_b - \tau_a)(1 + n \frac{1 - \alpha}{\alpha})} \sum_j Y_j$$
and a social planner can always increase welfare $W$ by lowering the tax rate $\tau_b$ toward $\tau_b^*$, regardless of the income distribution.

4 Empirics

We now examine nations’ tax revenues and bank deposits empirically, using panel data. We first report basic results based on simple regression models of the tax-to-GDP ratio, then consider some augmented models with controls for other determinants of tax. In each case the tax-bank link is economically and statistically significant, consistent with the theory developed in Section 3.

4.1 Basic Results

Our theoretical discussion focused on the link (3.2) between tax revenue and household spending via banks, each expressed as shares of output. Empirically, there is data on tax revenues for many countries, over some decades, but no corresponding data on spending via banks. There is however data on bank deposits. Let $D$ be the ratio of bank deposits to GDP, and let $\lambda = B/D$ be the ratio of bank spending to bank deposits, in which case $\lambda$ is the “velocity” of deposits as a form of money. Then we can rephrase the tax-bank link (3.2) as follows:

$$T = \tau_b \lambda D$$

(4.1)

If the marginal tax rate $\tau_b$ and deposit velocity $\lambda$ are fixed across countries and across time then (4.1) is a linear relationship between tax and bank deposits, each expressed as shares of GDP.

To examine the link between tax and bank deposits empirically, we use a panel dataset
covering 116 countries for which we have data on tax-to-GDP and deposits-to-GDP for some or all of the years in the range 1990-2008.\(^6\) Table 2 contains variable definitions and data sources, and Table 3 lists the countries – by region and level of economic development.

We show in the upper-left panel of Figure ?? a scatter plot of the tax and bank deposit variables, labelled TAX and BANK, based on the panel data described in Table 2. It is difficult to discern a clear historical relationship from the plot, but higher BANK seems to suggest higher TAX, on average. If so then the expectation \(E[TAX|BANK]\) should be increasing in BANK. To explore this possibility, in Table 4 (“levels” row, “panel data” columns) we report the ordinary least squares (OLS) coefficients for the simple linear regression model of TAX on BANK:

\[
TAX_{it} = \alpha + \beta BANK_{it} + u_{it}
\] (4.2)

for countries \(i = 1, \ldots, 116\) and years \(t = 1990, \ldots, 2008\), where \(u_{it}\) is the regression error – assumed independent of \(BANK_{it}\). We also report the regression \(R^2\) square, and the (pooled, first-order) autocorrelation between error \(u_{it}\) and it’s first temporal lag \(u_{i,t-1}\). As the residual autocorrelation is substantial, we explicitly model \(u\) as autoregressive:\(^7\)

\[
u_{it} = \phi u_{i,t-1} + \varepsilon_{it}
\] (4.3)

with \(\varepsilon_{it}\) serially independent and identically distributed over \(i\) and \(t\). Given the specification (4.2-4.3), and assuming further that errors \(u_{it}\) are independent across countries \(i\), for the TAX regression we report in Table 4 asymptotically valid standard errors.\(^8\) The result suggests a

---

\(^6\)Year 1990 is the earliest for which we have tax-to-GDP data.

\(^7\)The residual autocorrelations exhibit decay at longer lags, and for simplicity we model this as first-order autoregressive. The first-lag autocorrelations, as shown in Table 4, are (0.657, 0.649; 0.569, 0.570; -0.072, -0.070), while those at the second-lag are (0.435, 0.418; 0.376, 0.344; -0.136, -0.137), and those at the third-lag are (0.333, 0.313; 0.281, 0.270; -0.010, -0.012).

\(^8\)Our panel dataset has missing years \(t\) for some countries \(i\). We derive in Appendix A2 (Proposition 7) the relevant variance/covariance matrix for the OLS regression estimator, and apply the result via a
positive and statistically significant relationship between TAX and BANK.

Arguably, the regression errors $u_{it}$ in the TAX model (4.2) are correlated across countries, which may bias our efforts at inference. To address this possibility, in the setting of (4.3) suppose now that the disturbance terms $\varepsilon_{it}$ share a common component $\eta_t$, as follows:

$$\varepsilon_{it} = \eta_t + \omega_{it}$$

(4.4)

with $\omega_{it}$ a collection of mutually independent and identically distributed variables, each having mean zero and variance $\sigma^2_{\omega}$. To remove the influence of $\eta_t$ on inference, we subtract from each variable ($\text{TAX}_{it}$, $\text{BANK}_{it}$) its cross-section average at date $t$, resulting in transformed series – each a cross-section residual or contrast. We show in Figure ?? (top right) a scatterplot of the contrast series, and report in Table 4 ("levels" row, "panel contrast" columns) linear regression results for the transformed variables, analogous to those described earlier. As earlier the results suggest a positive and statistically significant TAX-BANK link.

We have interpreted the theoretical relationship (4.1) as one in which the marginal tax rate $\tau_b$ and deposit velocity $\lambda$ and constant across time and countries, but clearly both may vary. Earlier we modelled $\tau_b$ as endogenous, and the literature on monetary economics suggests that money velocity is endogenous as well. The expected tax, given bank deposits, takes the form of a product:

$$E[\text{TAX}|\text{BANK}] = E[\tau_b \lambda|\text{BANK}] \times \text{BANK}$$

when (4.1) holds, so if $\tau_b$ and/or $\lambda$ vary with BANK then expected tax is quite possibly non-linear in BANK. To address this point we take logarithms of the TAX and BANK variables, and show in Figure ?? (middle row) and Table 4 ("log-levels" row) the results for the log series, analogous to those for the level series. The scatter plots appear consistent with a

program GilbertIlievskiBasicResults.prg we wrote in EViews and which is available on the corresponding author’s website.

9See Friedman (1959) and Driscoll and Lahiri (1983).
linear relationship, the model fit is somewhat better in logs, and the results again suggest a positive and significant relationship between tax revenue and bank deposits.

The results so far are consistent with the co-emergence of the banking and public sectors, along the course of economic development. However, in the short-run there are other reasons that TAX and BANK might be linked. In a recession tax revenues may fall, as may money demand and bank deposits. If they fall in absolute terms and also relative to output then TAX and BANK may dip together in a recession. Attempts by the government to attenuate the business cycle may further contribute to the fluctuation of TAX and BANK short-term. Given these facts, the annual growth rates of TAX and BANK may be influenced more by short-term influences than by any co-emergence of banking and tax. For growth rates defined as first-differenced log-levels of TAX and BANK, we show in Figure ?? (bottom row) and Table 4 (“dif. log-levels” row) results analogous to those above. The results again suggest a positive and statistically significant link between tax revenue and bank deposits, but the $R^2$ squares are very low – consistent with the scatter plots.

The six models reported in Table 4 each address the idea that banks and taxation may emerge together within a given country. We need not interpret these results as causative, as the economic theory in Section 3 is one where tax and banking are jointly driven by changes in tastes, technology, etc. In terms of inference, we adjusted for the possibility of regression error correlations – across time and countries. Another approach is maximum likelihood estimation of a random effects model. For the three regressions in the “panel data” columns of Table 4, we get similar results when we estimate them via (country and time) random effects. Another possibility is fixed effects, but in that case the expectation of interest is no longer $E[TAX_{it}|BANK_{it}]$ but rather $E[TAX_{it}|BANK_{it}, \alpha_i, \mu_t]$, with fixed effects $\alpha_i$ and $\mu_t$. This is appropriate if one is testing whether bank changes cause tax changes, but for that we must step beyond the theoretical mindset of Section 3.
4.2 Extended Model

Suppose now that tax and banking outcomes are not purely the result of rational choice by utility-maximizing households and a well-meaning government, but instead reflect imperfections or outright failures in the banking sector. We do not attempt here a theoretical model of this situation, as this would require a treatment of corruption, industrial organization, and regulatory failure that far exceeds the present scope. Yet, if in fact the amount of spending via banks is somehow curtailed by failures in the banking system, in the framework of Section 3 (equation 3.1) tax revenues would also be curtailed. In other words, poorly functioning banks may form a bottleneck for the public sector’s funding.

To see whether bank deposits (relative to GDP) may cause tax revenues, by way of a bottleneck effect, we start by re-estimating the simple regression model (4.2) of TAX on BANK in log-log form\(^{10}\) via instrumental variables (to control for possible endogeneity of BANK), with the lagged BANK instrumenting for BANK itself, and with the inclusion of time and country fixed effects. The coefficient estimate for BANK is 0.37, the same as the OLS value reported in Table 4, with standard error now 0.014 rather than 0.010 as before. These results are consistent with the idea that BANK may cause variation in TAX.

We next augment the simple model (4.2) as follows:

\[
TAX_{it} = \alpha_i + \mu_t + \beta \text{BANK}_{it} + \gamma \text{BANK}_{it} \times \text{GDP}_{it} + \delta X_{it} + u_{it} \tag{4.5}
\]

with TAX, BANK and (per capita) GDP in log form, and an interaction effect BANK \times GDP that accommodates the possibility that the bottleneck effect of BANK on TAX may dissipate at higher levels of economic development. We also include a collection \(X\) of additional control variables, as well as fixed effects \(\alpha_i\) and \(\mu_t\). The idiosyncratic errors \(u_{it}\) are assumed independent and identically distributed across time and countries.

\(^{10}\)That is, we regress log-tax on log-deposits.
For controls we draw from the relevant tax literature, specifically Baunsgaard and Keen (2010), Auriol and Wanglers (2005), Rodrik (1998), and Tanzi (1987). The control variables are GDP per capita, trade openness, inflation, freedom\(^\text{11}\), aid per capita, population density, and lagged TAX. See Table 2 for data definitions and sources.\(^\text{12}\)

We estimate the model with fixed effects via the least squares dummy variable (LSDV) estimator.\(^\text{13}\) As the sample design has more countries than time periods, and the model includes a lagged endogenous variable, LSDV may exhibit important simultaneity bias. For robustness we follow up by instrumenting via the generalized method of moments (GMM), using difference GMM and system GMM panel methods,\(^\text{14}\) and report the results in Table 5.

In each column of Table 5, the BANK coefficient is positive and statistically significant. Also, the BANK-GDP interaction term is negative and significant in two out of three implementations of the model, consistent with the idea that the effect of banking on tax revenue is more prominent in lower income countries. For the control variables, coefficients are mostly insignificant except for lagged TAX.\(^\text{15}\)

Both the LSDV and GMM econometric approaches are possibly biased for estimating tax effects the model (4.5), with LSDV bias arising from endogeneity of lagged tax, and GMM bias occurring if some of the relevant moment conditions are violated. As a check

\(^{11}\)With freedom measured by the Civil Liberties score (Freedom House), a lower score indicates more liberty.

\(^{12}\)Of the 116 countries listed in Table 3, 5 of these have no data on some of our control variables, so the sample is a bit smaller here than in Section 4.1.

\(^{13}\)Hausman tests here favor fixed effects over random effects, so we do not report the latter.

\(^{14}\)For GMM we use Stata 11 software. Difference GMM is via Stata’s xtabond procedure, specified with no predetermined variables and no endogenous variables other than tax, with a maximum of 2 lags of tax used in the instrument set. Also, GMM is the two-step version, and standard errors are robust. System GMM is via the xtabond2 procedure, with the regression equation applied in levels and differences, with a single lag of differenced tax included as instrument in the levels equation, and with the GMM specification otherwise analogous to xtabond.

\(^{15}\)The GDP coefficient is positive but statistically insignificant in two out of three cases. Other studies have found similar results, and sometimes negative coefficients – see Rodrik (1998) and Baunsgaard and Keen (2010).
we compute from Table 5 Hausman tests for differences in coefficient estimates obtained from LSDV and GMM.\textsuperscript{16} For the coefficients of interest (BANK and BANK $\times$ GDP), the Hausman tests fail to reject the null of no LSDV bias at conventional significance levels. On the other hand, Sargan tests of GMM (overidentifying) moments reject the null of correct specification.\textsuperscript{17} Hence the fixed effects (LSDV) approach fares better.

The results are consistent with the idea that an increase in BANK may cause an increase in TAX, particularly at lower levels of GDP per capita, though for countries with well-functioning banks we need not interpret the bank-tax relationship as causative – but instead as co-emergent in the sense articulated in Section 3. At the least, the additional empirics strengthen the view that BANK and TAX are positively related.

\section{Conclusion}

The empirical evidence supports the hypothesis of co-emergence of banking and public sectors, along the course of economic development. As detailed in our theoretical discussion, there are various paths along which such co-emergence may happen. As a nation develops, its people may acquire more taste for public goods, motivating them to participate more in taxable activity, facilitating taxation and public good provision. Or, a shift in technology may increase the efficacy of transacting business via banks rather than cash, with similar tax effects. For the 116 countries we study empirically, which of these paths is more compelling? The answer is perhaps that each country follows its own development path that is influenced by shifts in tastes and technology.

For the U.S., tastes expressed early on by its leaders were for small government, particularly at the federal level, a sentiment that shifted dramatically by the 1930s. Data on the

\textsuperscript{16}That is, for each coefficient we compute the difference in estimates across methods, then square the difference, then divide the result by the difference in the coefficient’s GMM and LSDV squared standard errors, and compare to a $\chi^2$ critical value.

\textsuperscript{17}Sargan $p$ values are 0.041 and 0.000 for GMM-diff and GMM-sys, respectively.
historical development of the banking industry, the usage of banks for transactions, and the funding of public works, is perhaps as complete for the U.S. as for any other country. Hence future research could usefully trace the co-emergence of U.S. banking and public sectors.

Key to our theoretical discussion of tax and banks is the marginal tax rate – which for simplicity we take to be constant across sectors and income brackets. While a constant marginal tax rate is a fiction, in some countries this rate might be proxied by the value-added tax rate (VAT). Historical data on VAT is available for OECD countries, for example. In our theoretical framework, an endogenous tax rate responds to shifts in tastes and technology, as do the size of banking and public sectors. Future research can therefore examine the extent to which marginal tax rates move together with the tax-to-GDP and deposits-to-GDP ratios, along the course of economic development.

Our thesis is that the public sector co-emerges with the financial sector - measured by banks. More broadly, other components of the financial sector may be important for the government’s ability to generate tax. Formal markets for bonds and stocks may add further transparency to economic activity, and so may facilitate tax enforcement and collection. Additional work is needed to make this idea clearer in theoretical terms, and to test it empirically.

Acknowledgements: For comments on this work we thank participants at the Midwest Economic Association conference (spring 2011, St. Louis, Missouri), and the SIUC Economics Brown Bag seminar series (fall 2010 and 2011).
Appendix A1: Proofs of Economic Propositions

Proof of Proposition 1

Proof. The $i$-th agent maximizes utility by selecting spending levels $S_{ia}$ and $S_{ib}$ that provide the best consumption possibility. Substituting $S_{ia} = Y - S_{ib}$, and rearranging terms yields:

\begin{align*}
U_i &= C_1^{1-\alpha}C_2^{\alpha} \\
&= ((1 - \tau_a)S_{ia} + (1 - \tau_b)S_{ib})^{1-\alpha} \left( \tau_b \sum_{j=1}^{n} S_{jb} \right)^{\alpha} \quad (A-2) \\
&= ((1 - \tau_a)Y - (\tau_b - \tau_a)S_{ib})^{1-\alpha} \left( \tau_b \sum_{j=1}^{n} S_{jb} \right)^{\alpha} \quad (A-3)
\end{align*}

Differentiating with respect to $S_{ib}$, and setting the derivative equal to zero, yields:

\begin{equation}
S_{ib} + \frac{S_{ib} \left( \frac{1-\alpha}{\alpha} (\tau_b - \tau_a) - (1 - \tau_b) \right) + \frac{1-\alpha}{\alpha} (\tau_b - \tau_a) \sum_{j \neq i} S_{jb}}{1 - \tau_a} = Y \quad (A-4)
\end{equation}

Recast in terms of the $n \times 1$ vector $(S_{1b}, ..., S_{nb})'$, (A-4) yields:

\begin{equation}
A(S_{1b}, ..., S_{nb})' = b \quad (A-5)
\end{equation}

with $A$ an $n \times n$ matrix having elements:

\begin{align*}
A_{ii} &= \frac{(1 + \frac{1-\alpha}{\alpha})(\tau_b - \tau_a)}{1 - \tau_a} \quad (A-6) \\
A_{ij} &= \frac{\frac{1-\alpha}{\alpha}(\tau_b - \tau_a)}{1 - \tau_a}, \quad i \neq j \quad (A-7)
\end{align*}

and $b$ an $n \times 1$ vector with typical element $b_i = Y$. The solution is then $(S_{1b}, ..., S_{nb})' = A^{-1}b$, provided that the matrix inverse $A^{-1}$ exists. With $A$ having equal diagonal elements – all positive, and equal off-diagonal elements – all positive and less than the diagonal value, let $a$ be the diagonal value and $b$ be the off-diagonal value, with $0 < b < a$. Positing the inverse
$A^{-1}$ to have equal diagonal values, each taking value $c$, and equal off-diagonal values, each equal to $d$, the identity $AA^{-1}$ yields a system of equations in $c$ and $d$ with solution:

\[
\begin{aligned}
    d &= \left( (n - 1)b - \frac{a(a + (n - 2)b)}{b} \right)^{-1} \quad \text{(A-8)} \\
    c &= -\frac{d}{b} \frac{a + (n - 2)b}{b} \quad \text{(A-9)}
\end{aligned}
\]

which is well-defined because $(n - 1)b - \frac{a(a + (n - 2)b)}{b} < (n - 1)b - \frac{b(b + (n - 2)b)}{b} = 0$. The Nash equilibrium vector $(S_{1b}, \ldots, S_{nb})' = A^{-1}b$ therefore exists, is unique, and has the same value $S_{ib}$ for each household $i$.

As $S_{ib}$ is the same for each $i$, plugging them all into (A-4) and simplifying yields $S_{ib} = \frac{1}{(\tau_b - \tau_a)(1 + n\frac{1 - \alpha}{\alpha})} Y$, hence equation 3.7 holds. Given this fact, the remaining claims in the Proposition are straightforward to verify.

\[\square\]

**Proof of Proposition 3**

**Proof.** Evaluating CES utility given available choices yields:

\[
\begin{aligned}
    U_i &= \left( ((1 - \alpha)C_{i1}^\theta + \alpha C_{i2}^\theta \right)^{1/\theta} \\
    &= \left( (1 - \alpha)((1 - \tau_a)S_{ia} + (1 - \tau_b)S_{ib})^\theta + \alpha(\tau_b \sum_j S_{jb})^\theta \right)^{1/\theta} \quad \text{(A-11)} \\
    &= \left( (1 - \alpha)((1 - \tau_a)Y + (\tau_a - \tau_b)S_{ib})^\theta + \alpha(\tau_b \sum_j S_{jb})^\theta \right)^{1/\theta} \quad \text{(A-12)}
\end{aligned}
\]

where in (A-12) we use the fact that $S_{ia} = Y - S_{ib}$. Differentiating $U_i$ with respect to $S_{ib}$, and setting the derivative equal to zero yields:
\[(1 - \alpha)(\tau_a - \tau_b)((1 - \tau_a)Y + (\tau_a - \tau_b)S_{ib})^{\theta - 1} + \alpha \tau_b^\theta (S_{ib} + \sum_{j \neq i} S_{jb})^{\theta - 1} = 0 \quad (A-13)\]

Equivalently:

\[((1 - \alpha)(\tau_b - \tau_a))^{\frac{1}{\theta - 1}} ((1 - \tau_a)Y - (\tau_b - \tau_a)S_{ib}) = (\alpha \tau_b^\theta)^{\frac{1}{\theta - 1}} (S_{ib} + \sum_{j \neq i} S_{jb}) \quad (A-14)\]

This is a set of \(n\) linear equations and \(n\) unknowns \(S_{1b}, ..., S_{nb}\), generalizing the Cobb-Douglas case (A-4). Arguing as earlier, in matrix form this system is \(A(S_{1b}, ..., S_{nb})' = b\), with \(n \times n\) matrix \(A\) and \(b\) a vector with typical element \(b_i = Y\). Here the value \(a\) of \(A\)'s diagonal elements, and the value \(b\) of the off-diagonal elements, are:

\[a = (1 - \tau_a)^{-1} \left( \left( \frac{\alpha \tau_b^\theta}{1 - \alpha(\tau_b - \tau_a)} \right)^{\frac{1}{\theta - 1}} + (\tau_b - \tau_a) \right) \quad (A-15)\]

\[b = (1 - \tau_a)^{-1} \left( \left( \frac{\alpha \tau_b^\theta}{1 - \alpha(\tau_b - \tau_a)} \right)^{\frac{1}{\theta - 1}} \right) \quad (A-16)\]

with both \(a\) and \(b\) positive, and \(a > b\). The solution \((S_{1b}, ..., S_{nb})' = A^{-1}b\) therefore exists, with equal values \(S_{ib}\) for each household. Solving (A-14) for \(S_{ib}\) yields:

\[S_{ib} = \frac{(1 - \tau_a)((1 - \alpha)(\tau_b - \tau_a))^{\frac{1}{\theta - 1}}}{n \left( \alpha \tau_b^\theta \right)^{\frac{1}{\theta - 1}} + (\tau_b - \tau_a)((1 - \alpha)(\tau_b - \tau_a))^{\frac{1}{\theta - 1}}} Y \quad (A-17)\]

Simplification then yields equation 3.22 in the text. To sign the effects of \(n, \alpha, \tau_a, \tau_b\) on \(S_b\), the first is obvious, the second follows from the fact that \(\eta > 0\), and the remaining two are:
\[
\frac{\partial S_b}{\partial \tau_a} = -\frac{Y}{\tau_b - \tau_a + n\tau_b \left(\frac{1-\alpha}{\tau_a}\right)\eta} \quad (A-18)
\]

\[
-(1 - \tau_a)Y \frac{-1 + n\tau_b \eta \left(\frac{1-\alpha}{\tau_a}\right)\left(\frac{\tau_b - \tau_a}{\tau_a}\right)^{\eta-1}\tau_a^{-1}}{(\tau_b - \tau_a + n\tau_b \left(\frac{1-\alpha}{\tau_a}\right)\eta)^2} < 0
\]

\[
\frac{\partial S_b}{\partial \tau_b} = (1 - \tau_a)Y \frac{1 + n\left(\frac{1-\alpha}{\tau_a}\right)\left(\frac{\tau_b - \tau_a}{\tau_a}\right)^{\eta} + \tau_b \left(\frac{\tau_b - \tau_a}{\tau_a}\right)^{\eta-1}\tau_a^{-1}}{(\tau_b - \tau_a + n\tau_b \left(\frac{1-\alpha}{\tau_a}\right)\eta)^2} > 0 \quad (A-19)
\]

**Proof of Proposition 4**

*Proof.* Generalizing (A-12), utility is now:

\[
U_i = \left((1 - \alpha)\lambda_1^\theta((1 - \tau_a)Y + (\tau_a - \tau_b)S_{ib})^\theta + \alpha\lambda_2^\theta(\tau_b \sum_j S_{jb})^\theta\right)^{1/\theta} \quad (A-20)
\]

and proceeding as in (A-13)-(A-14) we get the desired formula for bank spending \(S_{ib}\). With \(\eta, \lambda_1\) and \(\lambda_2\) each positive numbers, the effects of \(n\) and \(\alpha\) on \(S_b\) are straightforward, and the effects of \(\tau_a, \tau_b\) on \(S_b\) are similar to those demonstrated in (A-18)-(A-18). The first is obvious, the second follows from the fact that \(\eta > 0\),

**Proof of Proposition 5**

*Proof.* Let the \(i\)-th household have income \(Y_i\). Then, in Nash equilibrium with spending via both cash and checks, check spending levels are:

\[
S_{ib} = cY_i + d \sum_{j \neq i} Y_j \quad (A-21)
\]
for constants $c > 0$ and $d < 0$ defined in the Appendix (equations A-9 and A-8), each a function of underlying parameters $\alpha, \tau_\alpha, \tau_b, n$. Solving for $c$ and $d$, in terms of the parameters, is a straightforward but lengthy exercise, yielding equations 3.27 and 3.28 in the text.

Aggregating over households $i$ yields total check spending:

$$\sum_i S_{ib} = (c + (n - 1)d) \sum_i Y_i$$  \hspace{1cm} (A-22)

Since (A-22) must hold when income is homogeneous ($Y_1 = \cdots = Y_n$), we can evaluate the multiplier $c + (n - 1)d$ in that case via Proposition 1.

\[ \square \]

**Appendix A2: Panel Econometrics**

Consider the linear model:

$$y_{it} = x_{it}\theta + u_{it}$$  \hspace{1cm} (A-23)

for a panel of data with dependent variable $y_{it}$, a $1 \times K$ vector $x_{it}$ consisting of independent variables and possibly a constant/intercept term, and a zero-mean finite-variance error $u_{it}$ which is independent of $x_{it}$ but may be correlated across dates $t$. Suppose also that the panel may be unbalanced, and let $T_i$ be the number of dates available for the $i$-th unit.

Let the model (A-23) be estimated via OLS, with coefficient estimator $\hat{\theta}$. Correlation among errors $u_{it}$ can effect the sampling distribution of $\hat{\theta}$, including the variance/covariance matrix $V(\hat{\theta})$ of the OLS estimator. The following proposition derives this covariance matrix when errors are first-order autocorrelated across dates $t$ but independent across units $i$, in a possibly unbalanced panel:
Proposition 7. For a possibly unbalanced panel of data, suppose that the linear model (A-23) holds with errors $u_{it}$ independent across cross-section units $i$ but first-order stationary autoregressive across time $t$. Then the OLS coefficient vector $\hat{\theta}$ has a variance/covariance matrix $V(\hat{\theta})$ which, conditional on the $x$ data, has the following form:

$$V(\hat{\theta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1}$$  \hspace{1cm} (A-24)

where $X$ is the $K$-column matrix of stacked data $x_{it}$ sorted by $i$ then $t$, $u$ is the vector of stacked errors, and:

$$E(uu') = \begin{pmatrix}
\Lambda_1\Gamma\Lambda_1' & 0 & \cdots & 0 \\
0 & \Lambda_2\Gamma\Lambda_2' & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & \Lambda_n\Gamma\Lambda_n'
\end{pmatrix}$$  \hspace{1cm} (A-25)

where $0$ indicates a matrix of zeros, $\Gamma$ is the $T \times T$ matrix with typical element $\Gamma_{jk} = \sigma^2 \phi^{|j-k|}$, and $\Lambda_i$ is the $T_i \times T$ matrix for which $\Lambda_{ijk} = 1$ at row $j$ and column $k$ if $j$ is the $j$-th record of available data (sorted by year) and $k$ is the year for that record, and for which $\Lambda_{ijk} = 0$ otherwise.

Proof. The variance form (A-24) is standard for OLS in the presence of error correlation (but no conditional heteroskedasticity), see for example Ruud (2000, equation 18.4). For a balanced panel, with errors first-order autocorrelated over time, the term $E(uu')$ in A-24 is:

$$E(uu') = I_n \otimes \Gamma$$  \hspace{1cm} (A-26)

with $I_n$ the $n \times n$ identity matrix. The situation for an unbalanced panel is complicated by the fact that the same autocovariance matrix $\Gamma$ can not be directly applied to the errors $u_{it}$ for each member $i$, as some members have missing time observations. Proceeding indirectly,
the matrix $\Lambda_i$ serves to indicate the non-missing values for member $i$, allowing us to write the covariance matrix of the $i$-th member’s errors as $\Lambda_i \Gamma \Lambda_i'$, resulting in (A-25). Note that for a balanced panel, $\Lambda_i = I_T$ for each $i$ and (A-25) reduces to (A-26).

□

To apply the variance formula (A-24) we first compute OLS residuals $\hat{u}_{it}$, regress residual on lagged residual via OLS, then plug in the estimates $\hat{\phi}$ and $s^2_\varepsilon$ into (A-24)-(A-25). For this we have written an EViews computer program GilbertIlievskiBasicResults.prg, available on the corresponding author’s website.
References


Table 1: Average values of government tax revenue and bank deposits as a share of GDP.

<table>
<thead>
<tr>
<th>Economic Level</th>
<th>GDP per capita</th>
<th>Tax (% of GDP)</th>
<th>Bank Deposits (% of GDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>1005 and less</td>
<td>11.93</td>
<td>18.58</td>
</tr>
<tr>
<td>middle</td>
<td>1006 - 3975</td>
<td>14.89</td>
<td>30.01</td>
</tr>
<tr>
<td></td>
<td>3976 - 12275</td>
<td>18.72</td>
<td>40.52</td>
</tr>
<tr>
<td>high</td>
<td>12276 and higher</td>
<td>31.66</td>
<td>88.07</td>
</tr>
</tbody>
</table>

Table 2: Variable Description

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>gross domestic product per capita in constant (year 2000) dollars, adjusted for PPP. Source: online version of the World Development Indicators (WDI).</td>
</tr>
<tr>
<td>TAX</td>
<td>Tax revenue divided by GDP. This ratio downloaded from the Government Finance Statistics. For OECD countries, missing values obtained from the OECD database online.</td>
</tr>
<tr>
<td>BANK</td>
<td>Deposits divided by GDP, with deposits being demand, time, and saving deposits. This ratio obtained from the electronic version of International Financial Statistics (IMF), October 2008.</td>
</tr>
<tr>
<td>Aid</td>
<td>The amount of official development assistance (grants plus concessional loans, measured in U.S. dollars) divided by Gross National Income. Source: WDI.</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Share of agriculture in aggregate value added. Source: WDI.</td>
</tr>
<tr>
<td>Density of Population</td>
<td>The midyear population divided by land area in square kilometers. Source: WDI.</td>
</tr>
<tr>
<td>Inflation</td>
<td>Growth rate of the consumer price index. Source: WDI.</td>
</tr>
<tr>
<td>Openness to Trade</td>
<td>The sum of exports and imports of goods and services measured as a share of GDP. Source: WDI.</td>
</tr>
<tr>
<td>Region</td>
<td>Income Level</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Africa</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>middle</td>
</tr>
<tr>
<td>Americas</td>
<td>middle</td>
</tr>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Asia &amp; Pacific</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>middle</td>
</tr>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Middle East</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>middle</td>
</tr>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Post-Soviet</td>
<td>middle</td>
</tr>
<tr>
<td></td>
<td>high</td>
</tr>
<tr>
<td>Western Europe</td>
<td>high</td>
</tr>
</tbody>
</table>
Table 4: Regressions of Tax Revenue on Bank Deposits

<table>
<thead>
<tr>
<th>series type</th>
<th>panel data</th>
<th>panel contrasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>intercept</td>
<td>slope $R^2$</td>
</tr>
<tr>
<td>levels</td>
<td>16.616</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(0.429)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>log-levels</td>
<td>1.592</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>dif. log-levels</td>
<td>0.009</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.007)</td>
</tr>
</tbody>
</table>

Table 5: Multivariate Regressions of Tax Revenue

<table>
<thead>
<tr>
<th></th>
<th>Fixed Effects</th>
<th>Difference GMM</th>
<th>System GMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANK</td>
<td>0.34</td>
<td>0.79</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>(0.142)**</td>
<td>(0.391)**</td>
<td>(0.296)**</td>
</tr>
<tr>
<td>BANK × GDP</td>
<td>-0.03</td>
<td>-0.08</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.016)**</td>
<td>(0.040)**</td>
<td>(0.033)</td>
</tr>
<tr>
<td>GDP</td>
<td>0.12</td>
<td>0.4</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.186)**</td>
<td>(0.151)</td>
</tr>
<tr>
<td>lag TAX</td>
<td>0.68</td>
<td>0.43</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(0.040)*****</td>
<td>(0.145)*****</td>
<td>(0.151)*****</td>
</tr>
<tr>
<td>Agriculture</td>
<td>-0.003</td>
<td>-0.01</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)**</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Aid</td>
<td>-0.001</td>
<td>-0.004</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)*</td>
<td>(0.002)**</td>
</tr>
<tr>
<td>Density</td>
<td>-0.002</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.001)**</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Inflation</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Liberty</td>
<td>-0.004</td>
<td>0.002</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Trade</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, *** (**, *) indicate significance at 1 (5, 10).