Are Theory-Based Debt Sustainability Indicators Useful for Predicting Crises?

Katharina Mersmann[1]
Osnabrück University

Frank Westermann[2]
Osnabrück University and CESifo

Abstract

A large literature in empirical public finance applies time series techniques to historical data and draws inference about public debt sustainability of individual countries. These methods include unit root tests on primary deficits, co-integration between revenue and expenditure as well as fiscal reaction functions. In this paper, we take a systematic approach to evaluating the in- and out-of-sample performance of various methods in predicting sovereign debt crises. In a panel-logit regression analysis, we find only very limited benefits for forecasting.

JEL Classification Numbers: H6, E62

Keywords: intertemporal budget constraint, unit roots, co-integration, fiscal reaction function

*We thank M. Hutchison, K. Kletzer, and M. Melvin for helpful comments and discussions.

[1] Institute of Empirical Economic Research, Osnabrueck University, Rolandstraße 8, 49074 Osnabrück. Email: katharina.mersmann@uni-osnabrueck.de
[2] Institute of Empirical Economic Research, Osnabrueck University, Rolandstraße 8, 49074 Osnabrück. Email: frank.westermann@uni-osnabrueck.de

1 Introduction

Public debt sustainability is one of the major concerns of policymakers, in particular since the global financial crisis in 2007/8. Researchers have used a variety of empirical approaches to test for sustainability, and have applied these tests to individual countries. Nevertheless, there is no systemic evidence that analyses the link between sustainability and the incidence of sovereign debt crises – a gap in the literature that this paper intents to fill.

We consider three main approaches that have been used to assess the sustainability of public debt: The first group has focused on unit root tests for public debt or deficits (see e.g., Hamilton & Flavin, 1986). A second generation of indicators built on co-integration tests for debt and primary balances (Trehan & Walsh, 1988), or government revenue and spending (Hakkio & Rush, 1991). A third group is based on fiscal rection functions of current deficits to lagged values of government debt (see Bohn (1998, 2007)). In each case, the idea is to evaluate whether a certain path of expenditure and revenues is consistent with the intertemporal budget constraint.

We relate signals from these indicators (sustainable/unsustainable) to indices of sovereign debt crises, based on Gerling et al. (2017). The indices include a very narrow definition of defaults, as well as wider definitions, such as IMF borrowing, periods of high inflation, or the central bank stepping in as a lender of last resort. We also used the Reinhart & Rogoff (2011) domestic debt crisis index as a robustness test.

The main finding of our paper —independent of the signal or crisis definition used— is that there exists very little relationship between theory-based sustainability signals and the incidence of sovereign debt crisis. Both in- and out-of-sample, the indicators only marginally add information about the occurrence of crises. We establish this finding in a set of descriptive statistics and in a panel-logit regression analysis.

Our main result is reminiscent of earlier findings on twin banking- and currency crises, as well as exchange rate models. Meese & Rogoff (1983) found that none of the structural exchange rate models is able to predict the exchange rate out of sample, at least on short-term forecast horizons. Kaminsky & Reinhart (1999) have documented that theory-based leading indicators on currency crises add little to the forecasting performance. In our paper, we draw on their empirical approaches, in particular the noise-to-signal ratio, to evaluate the performance of various models.

Our findings do not challenge the theoretical correctness of models that were designed to test whether a given path of deficits and debt is consistent with the intertemporal budget. Instead, our results imply that sovereign defaults appear to occur for reasons other than violations of the intertemporal budget constraint. These other factors could for instance include political-economy considerations, a cost-

benefit analysis of international debt repayment, and historical factors (see Kletzer & Wright, 2000; Manasse & Roubini, 2009; Reinhart et al., 2003), the cyclical state of the business cycle (Tomz & Wright, 2007), or market sentiments and contagion (Aizenman et al., 2013; De Grauwe & Ji, 2013).

2 Data and Methodology

2.1 Data

The data on gross public debt (debt), primary balance (pb), government expenditures (exp) and revenues (rev), all expressed as a share of GDP, are taken from the IMF Historical Public Finance Dataset by Mauro et al. (2013). In order to have a reasonable cross section of countries, we start the analysis in 1950. The series are updated with the most recent data from the IMF's World Economic Outlook (WEO). The final dataset contains 31 countries with yearly observations from 1950 to 2015.[3] We applied the Hodrick–Prescott filter to decompose output and government spending into trend and cyclical components; the real output series are taken from the Total Economy Database (TED).[4]

Our baseline indicator for fiscal crises is taken from Gerling et al. (2017). Crises events are defined as a binary variable including any debt operation resulting in an economic loss on private creditors, including outright default, restructuring, or rescheduling events from 1970 to 2015. The indicator builds on the Bank of Canada's CRAG sovereign defaults database, which considers a particular year to be a default year, when "debt service is not paid on the due date or within a specified grace period, when payments are not made within the time frame specified under a guarantee, or, absent an outright payment default" (Beers & Mavalwalla, 2017, p. 2). This definition is consistent with the literature on sovereign defaults (e.g., Cruces & Trebesch, 2013), and the practices of credit-rating agencies.

We also used wider definitions of sovereign debt crises proposed by Gerling et al. (2017), discussed in more detail in the appendix. Furthermore, we considered the Reinhart and Rogoff (2011) indicator as a robustness test. This indicator includes both narrow and wide definitions of crises, but is limited to cases of debt obligations issued under domestic law.

[3] Countries in our sample are Argentina, Australia, Austria, Belgium, Canada, Colombia, Denmark, Finland, France, Germany, Greece, Honduras, Iceland, India, Ireland, Italy, Japan, Mexico, the Netherlands, New Zealand, Norway, Pakistan, Philippines, Portugal, Spain, Sweden, Switzerland, Thailand, the United Kingdom, the United States and Venezuela.

[4] Because data for Honduras are missing in TED, we compute the GDP series using Maddison data (MAD) before 1960 and for later years, the World Development Indicators (WDI) database.

2.2 Assessing fiscal sustainability

We assess the sustainability of public debt by computing seven different theory-based indicators often used in the literature:

First, we classify fiscal policy for each country unsustainable if a unit root is found in the debt, primary balance or first difference of debt series (d(debt)), using augmented Dickey and Fuller (1979), Phillips and Perron (1988), Kwiatkowski et al. (1992) and Zivot and Andrews (1992) unit root tests. If at least three of four test statistics indicate an unsustainable path, we classify this as a crisis signal.[5]

Second, we apply the Johansen (1995) co-integration test, with Osterwald and Lenum (1992) critical values. If co-integration is found between debt and primary balance, or between revenue and expenditure, we classify the country as sustainable. No co-integration leads to a crisis-signal.[6]

Finally, in the case of the fiscal reaction function, a positive and statistically significant sign of the lagged debt variable is considered to signal sustainability. Here we test the "simple" and extended versions suggested by Bohn (1998). In the extended version, we add determinants based on Barro's (1979) tax-smoothing model as control variables.[7]

3 In-Sample Performance

As a first step, and to anticipate the further results, we analyze the link between sustainability indicators and sovereign debt crisis over the full sample period, from 1950 to 2015. Table 1 shows the accuracy of predicting in-sample crises using all seven indicators of fiscal sustainability and six definitions of financial crisis. The reported values display the percentage of correct signals received.

Table 1 documents that most sustainability measures perform very poorly when predicting crises in-sample. 50% can be considered a benchmark here, as tossing a coin would also predict half of the crises correctly in a large enough sample. We find that most indices fail to reach this benchmark. This is the case for narrow definitions of crises, as well as for more broadly defined crises. We also use a joint indicator, which is equal to one if any crises indices 1 to 4 signal a crisis.

The unit root test on the change in debt performs somewhat better than the other indicators. In all cases, it has a signaling power above 50%. For some crises indices, the percentage is as high as 80% and 81%. Note, however, that these are unconditional probabilities. In the following sections, we analyze whether this above-average in-sample fit can be used to make out-of-sample predictions.

---

[5] All four tests are often used to assess sustainability in the literature. In the robustness section, we take into account different orderings among the tests.

[6] We consider countries as unsustainable, when one series is I(1) and the other is I(0). In this case, co-integration would be merely an artifact of placing all weight in the co-integration vector on one of the variables.

[7] Controls are the level of temporary government spending and a business cycle indicator (see Bohn, 1998). We follow Mendoza and Ostry (2008) and use the expenditure and output gaps derived from the HP-filter as proxies.

Table 1 Signaling power − Insample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Crises criterion			
Indicator	Joint	1	2	3	4	5	Domestic
Unit root (debt)	39%	29%	29%	39%	26%	29%	17%
Unit root (pb)	42%	52%	52%	42%	48%	52%	57%
Unit root d(debt)	55%	71%	71%	55%	81%	65%	80%
Cointegration (debt & pb)	55%	45%	39%	55%	35%	45%	30%
Cointegration (rev & exp)	45%	42%	42%	45%	26%	35%	27%
Fiscal reaction function	35%	32%	32%	35%	42%	45%	40%
Extended fiscal reaction function	39%	35%	35%	39%	45%	35%	43%

Notes: "debt"= Debt to GDP ratio; "pb"=Primary Balance; d(debt)= Change in debt; "rev & exp"=Revenues and Expenditure. The table displays the sum of correct signals as a share of total signals. For details on crises indicators and data see Appendix.

4 Out-of-Sample Forecasts

For out-of-sample forecasts, we exploit the full panel dimension of our data set. In Table 2, we report the descriptive statistics based on a rolling 35-year window. Several recent academic papers on debt sustainability, using panel data, have been published using a comparable sample size.[8] For each window, we estimate our seven indicators and see whether a crisis occurs in the following 5 year-window. Using this approach, we have 837 data points and 79 windows of 5 years in which crises are following.

We start by looking again at unconditional probabilities and Type I and Type II errors. Table 2 illustrates that Type II errors are often very large. From 16.13% (unit root test on the change in debt) to 85.42% (unit root test on debt) of the cases, we are receiving a crisis-signal, although no crisis occurred in the following 5 years. Furthermore, when sustainability signals were received, some countries experienced a sovereign debt crisis in the follow 5-year window (Type I error). The signaling power, which is the share of correct signals [1-(sum of Type I+Type II errors)], is below 50% for all indicators, except for the unit root test on the first difference in debt, and for the extended fiscal reaction function. The unit root test, which appears to perform quite well in terms of Type II errors, however also displays the largest share of Type I errors. Its relatively high signaling power thus largely derives from the ability to avoid "false alarms," but comes at the expense of frequently missing to signal crisis events when they happen.

[8] See e.g., Debrun and Kinda (2016; 30 years), Mauro et al. (2015; rolling 25 years), Fincke and Greiner (2012; 38 years).

Table 2: Signaling Power − Out-of-Sample

	(1)	(2)	(3)
Indicator	Correct Signals	Type I	Type II
Unit root (debt)	14%	0.36%	85.42%
Unit root (pb)	42%	3.23%	54.72%
Unit root d(debt)	76%	8.12%	16.13%
Cointegration (debt & pb)	31%	0.84%	68.34%
Cointegration (rev & exp)	25%	1.19%	73.84%
Fiscal reaction function	43%	3.58%	53.17%
Extended fiscal reaction function	52%	2.03%	46.36%

Notes: (1) Sum of correct signals as a share of total signals (i.e., 1- Type I -Type II). (2) Type I denotes the probability of a sustainability signal and at least one crisis 5 years ahead. (3) Type II is the probability of an unsustainability signal and no crisis 5 years ahead.

Table 3 takes a slightly different approach, by reporting conditional probabilities and the noise-to-signal-ratio proposed in Kaminsky and Reinhart (1999). The first two columns display the probability of receiving an *unsustainable* signal (not sustainable, "nS"), given that a crisis occurred, respectively that no crisis occurred. The ratio of the two—the "noise-to-signal" ratio—should be smaller than one to be useful for forecasting purposes. As Column (3) shows, this is the case for all but one indicator: the unit root test on the change in debt. Columns (4) and (5) show that for most indicators, the probability of observing a crisis, given a crisis signal, is larger than the unconditional probability of a crisis only by a very small margin. The signal from the extended fiscal reaction function raises the probability of actually observing a crisis in the following five years, after receiving the signal, by 4.34%. For one indicator, the signal actually lowers the probability of observing a crisis.

Table 3: Conditional Probabilities and Noise-to-Signal- Ratio

	(1)	(2)	(3)	(4)	(5)
Indicator	P(nS| crisis)	P(nS| no Crisis)	Noise-to-signal	P(crisis|nS)	P(crisis|nS) - P(crisis)
Unit root (debt)	96%	94%	0.98	9.61%	0.17%
Unit root (pb)	66%	60%	0.92	10.20%	0.76%
Unit root d(debt)	14%	18%	1.28	7.53%	-1.90%
Cointegration (debt & pb)	91%	75%	0.83	11.18%	1.74%
Cointegration (rev & exp)	87%	82%	0.93	10.04%	0.61%
Fiscal reaction function	62%	59%	0.95	9.92%	0.48%
Extended fiscal reaction function	78%	51%	0.65	13.78%	4.34%

Notes: (1) Unsustainable signal, conditional on a crisis in the following 5 years, P(nS|C); (2) Unsustainable signal, conditional on no-crisis in the following 5 years, P(nS|nC); (3) Noise to Signal: Ratio of false and true signal P(nS|nC) /P(nS|C); (4) Probability of a crisis, conditional on a unsustainable signal.

The positive impact of some indicators is rather small, however, from the perspective of a contingency analysis. The Pearson's chi-squared and Fisher's exact test, for instance, indicate that only two of seven indicators have a significant link to the occurrence of a crisis, at the 5% level (for co-integration between debt and primary balance, as well as the extended fiscal reaction function). Furthermore, Cramer's *V* indicates that the strength of the link is rather marginal, as in all cases the value is below

0.2[9]. When considering the Goodman and Kruskal´s τ statistic, we find that the reduction in the forecast error of predicting a crisis, given a signal of unsustainability, is only marginal. For all of our seven indicators we find a reduction in prediction error of less than 5%; in most cases even less than 1%.

5 Regression Analysis

5.1 A multivariate logit regression

In the last approach, we investigate whether the crisis indicators might jointly have predictive power for the occurrence of sovereign debt crisis. Table 4 reports the results of a set of logit regressions. First, we investigate each indicator individually in a binary regression of the crisis index on each of the theory-based sustainability indices. The result is that most sustainability indexes are statistically insignificant in this regression. The only exceptions are signals from the fiscal reaction function, which are statistically significant at the 5% and 10% level. Receiving a signal from this index raises the probability of a crisis in the following 5 years by about 3.45 pp and 2.87 pp.

Table 4: Logit Regression

	Crises 5 year ahead							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unit root (debt)	-0.0110							-0.0104
	(0.0321)							(0.0324)
Unit root (pb)		0.0199						0.0372**
		(0.0177)						(0.0168)
Unit root d(debt)			-0.0250					-0.0163
			(0.0251)					(0.0241)
Cointegration (debt & pb)				-0.0069				-0.0198
				(0.0212)				(0.0201)
Cointegration (rev & exp)					-0.0087			-0.0020
					(0.0192)			(0.0180)
Fiscal reaction function						0.0345**		0.0315
						(0.0160)		(0.0233)
Extended fiscalreaction function							0.0287*	0.0188
							(0.0169)	(0.0256)
Obs	837	837	837	837	837	837	837	837
Chi^2	0.1140	1.2365	0.9292	0.1035	0.2004	4.5638	2.7801	9.5732
p	0.7356	0.2661	0.3351	0.7477	0.6544	0.0327	0.0954	0.2141

Notes: Marginal effects: dy/dx for discrete change of dummy variable from 0 to 1; Standard errors are in parentheses;
*** denotes a significance level of 1%, ** of 5%, and * of 10 %.

When including all indicators, the primary balance turns statistically significant at a 5% significance

[9] For the extended fiscal reaction function, Cramers V is 0.16 and the Goodman and Kruskal's statistic is 0.0256.

level, whereas the fiscal reaction functions both turn insignificant. The marginal effect of the primary balance, however, is again quite small, raising the probability of observing a crisis only by 3.72 pp. Thus a multiple regression approach leaves the main conclusion, the general absence of predictive power of the signal, largely unchanged.

5.2 Robustness

We have tested the robustness of our findings in several different ways, focusing on whether the absence of forecasting power is due to our choice of sample length or particular definitions of crises, types of unit root tests, etc. We are able to slightly improve our results, but the overall assessment that the use of theory-based crisis indices is extremely limited for forecasting, remains highly robust.

Crises Indicators

In the benchmark regression, we use the most narrow definition, that is, actual private default on public debt. However, in a wider sense, a range of other indices signal a sovereign debt problem. As pointed out in 2.1 we use the five crises criteria by Gerling et al. (2017). The criteria range from an actual credit event, the actual use of official financing, implicit defaults measured by high inflation and domestic arrears accumulation, and loss of market confidence.[10] Furthermore we use the domestic debt crisis indicator by Reinhart and Rogoff (2011). None of these indicators has a better in- or out-of-sample forecasting performance than the one used in the benchmark regression.

Unit root tests

While in our benchmark regression, we take an agnostic approach by requiring at least three of four unit root tests to signal an unsustainable path, we also experimented with a certain hierarchy among the tests. We first considered the ADF and Phillips–Peron test only. If both signal a unit root, we moved on and implemented the KPSS test, which turns the null hypothesis around. If we continue to obtain a signal on unsustainability, we implemented the ZA-test, which allows for a structural break in the trend. Only if this test also cannot reject the unit root, we classified the period as unsustainable. This approach addresses the issue of a high share of Type II errors in our baseline findings. The robustness test has been designed to have a strong bias against sending false alarms, as periods are only classified as unsustainable when really strong evidence of a unit root exists. Nevertheless, only little improvement in the forecasting quality is observable: Although the Type II errors (false signals) diminish in this approach, Type I errors (crises without signals) increase.

[10] For a more detailed description, see Gerling et al. (2017).

Sample length

A final set of robustness tests is regarding the sample length. Here we faced the following trade-off: When choosing a longer sample, we have greater statistical power in the unit root and co-integration tests; when choosing a shorter sample, we have more out-of-sample observations to conduct descriptive statistics and logit-regressions. As a benchmark, we opted for a 35-year window, which is not uncommon in the recent literature, although some classical papers use sample sizes that are substantially longer. We tested a range of sample sizes from 30 to 40 years[11] and indeed observed the maximum forecasting power in a 40-year window, but it is again only marginally higher than in our benchmark regression.

6 Conclusions

In sum, we find very little evidence that theory-based sustainability indices help predict sovereign debt crises. This finding is kown from other areas of economics, such as exchange rate forecasting and predicting banking and currency crises. The finding is nevertheless surprising in light of the widespread use of these indicators in academic policy research, and the often plausible findings reported in earlier studies.

Our analysis should be interpreted as a cautious reminder that countries satisfying the usual conditions for debt sustainability may still be prone to sovereign debt crisis, and those that do not may be doing well without a crisis for many years in the future.

7 References

Aizenman, J., Hutchison, M., & Jinjarak, Y. (2013). What is the risk of European sovereign debt defaults? Fiscal space, CDS spreads and market pricing of risk. *Journal of International Money and Finance, 34*, 37–59.

Barro, R. J. (1979). On the determination of the public debt. *Journal of Political Economy, 87,* 940–971.

Beers, D. T., & Mavalwalla, J. (2017). *Database of Sovereign Defaults, 2017*. Retrieved from https://www.bankofcanada.ca/2014/02/technical-report-101/

Bohn, H. (1998). The behavior of US public debt and deficits. *Quarterly Journal of Economics, 113,* 949–963.

Bohn, H. (2007). Are stationarity and cointegration restrictions really necessary for the intertemporal budget constraint? *Journal of Monetary Economics, 54,* 1837–1847.

Cruces, J. J., & Trebesch, C. (2013). Sovereign defaults: The price of haircuts. *American Economic Journal: Macroeconomics, 5*(3), 85–117.

De Grauwe, P., & Ji, Y. (2013). Self-fulfilling crises in the Eurozone: An empirical test. *Journal of International Money and Finance, 34*, 15–36.

[11] We also checked up to the maximum range of 58 years. This exercise reduced our data set to a cross section and we tried to predict crises that occurred in the post-2008 period. As in the panel-data analysis, there is no evidence on the predictive power of the sustainability signals.

Debrun, X., & Kinda, T. (2016). That squeezing feeling: The interest burden and public debt stabilization. *International Finance, 19,* 147–178.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74,* 427–431.

Fincke, B., & Greiner, A. (2012). How to assess debt sustainability? Some theory and empirical evidence for selected euro area countries. *Applied Economics, 44,* 3717–3724.

Gerling, M. K., Medas, M. P. A., Poghosyan, M. T., Farah-Yacoub, J., & Xu, Y. (2017). *Fiscal crises*: Washington, DC: International Monetary Fund.

Hakkio, C. S., & Rush, M. (1991). Is the budget deficit "too large?". *Economic inquiry, 29,* 429–445.

Hamilton, J. D., & Flavin, M. A. (1986). On the limitations of government borrowing—A framework for empirical testing. *American Economic Review, 76,* 808–819.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford, England: Oxford University Press.

Kaminsky, G. L., & Reinhart, C. M. (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review, 89,* 473–500.

Kletzer, K. M., & Wright, B. D. (2000). Sovereign debt as intertemporal barter. *American Economic Review, 90,* 621–639.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54,* 159–178.

Manasse, P., & N. Roubini (2009). "Rules of thumb" for sovereign debt crises. *Journal of International Economics, 78*(2), 192–205.

Mauro, M. P., Romeu, R., Binder, M. A. J., & Zaman, M. A. (2013). *A modern history of fiscal prudence and profligacy* (Working paper No. 13/5). Washington, DC: International Monetary Fund.

Mauro, P., Romeu, R., Binder, A., & Zaman, A. (2015). A modern history of fiscal prudence and profligacy. *Journal of Monetary Economics, 76,* 55–70.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics, 14*(1–2), 3–24.

Mendoza, E. G., & Ostry, J. D. (2008). International evidence on fiscal solvency: Is fiscal policy "responsible"? *Journal of Monetary Economics, 55,* 1081–1093.

Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics, 54,* 461–472.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75,* 335–346.

Reinhart, C. M., & Rogoff, K. S. (2011). From financial crash to debt crisis. *American Economic Review, 101,* 1676–1706.

Reinhart, C. M., Rogoff, K. S., & Savastano, M. A. (2003). Debt intolerance.

Tomz, M., & Wright, M. L. J. (2007). Do countries default in "bad times"? *Journal of the European Economic Association, 5,* 352–360.

Trehan, B., & Walsh, C. E. (1988). Common trends, the government's budget constraint, and revenue smoothing. *Journal of Economic Dynamics and Control, 12,* 425–444.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics, 10,* 251–270.

Appendix: Definitions for crisis indicators

Table 5: Crises Indicators[12]

Variable	Description
Criterion (1)	Any operation that makes creditors incur material economic losses on the sovereign debt they hold (for instance, default, restructuring, rescheduling, or relief).
Criterion (2)	A subset of Criterion (1) that only considers losses of private investors
Criterion (3)	Years with exceptionally large official IMF financing, with fiscal adjustment as a program objective.
Criterion (4)	Implicit defaults through high inflation rates and steep outstanding domestic debt accumulation.
Criterion (5)	High prices of bond issuance (signaling loss of market confidence).
Criterion (Joint)	Captures any crises event defined by Criteria 1 to 4.
Criterion (Domestic)	All debt crises involving outright default on payment of debt obligations issued under domestic law.

[12] Crises Criterion 1 to 5 and Joint are from Gerling et al. (2017), whereas Criterion Domestic is from Reinhart & Rogoff (2011).