Aarhus and Hamburg,

September 6th - 2010

The transition of corruption

Causality and two models

Erich Gundlach, GIGA and Hamburg University, Germany[1]

Martin Paldam, School of Economics and Management, Aarhus University, Denmark[2]

Abstract:

Development is a grand process of transitions, where the level of many variables shifts from one level in low income countries to a different level in developed countries. The transition of corruption is from high corruption to high honesty. The agricultural, the demographic, and the democratic transitions are classical textbook examples. Recently, a new empirical method has been used to weed out any potential spuriousness in the observed transitions. Elsewhere we have shown that the new method finds that the three textbook transitions are spurious, and we have discussed what this means. The present note shows that the new method makes the transition of corruption go away as well.

1. Address: GIGA German Institute of Global and Area Studies/Institute of Asian Studies. Rothenbaumchaussee 32, 20148 Hamburg, Germany.
E-mail:. gundlach@giga-hamburg.de. Homepage: http://www.erichgundlach.de
2. Address: School of Economics and Management, Bartholins Allé 10, DK-8000 Aarhus C, Denmark.
E-mail: mpaldam@econ.au.dk. URL: http://www.martin.paldam.dk

1. Introduction

This note is meant as a background paper to Gundlach and Paldam (2010). It compares two causality tests as applied to transition processes in economic development. The long-run DPIV-test (from Gundlach and Paldam (2009a) is presented in a set of A-tables and the shorter run AJRY-test from Acemogly et al. (2008) is reported in a set of matching B-tables.

This note looks at the case of the transition of corruption: The level of corruption-/honesty changes from a high level of corruption in traditional low income countries to a high level of honesty in modern developed countries. We have studied the transition in other papers: Paldam (2002) showed that income is the best explanatory variable in cross-country regressions of corruption. Gundlach and Paldam (2009b) analyze long-run causality using the DPIV-test. As our study was published in *Economic Letters* a very condensed presentation was necessary, so the full A-table is reported below.

The estimation equations for the A- and the B-table are given in Table 1. They are carefully discussed in Gundlach and Paldam (2010). At present we do not repeat the argument. The note proceeds as follows. Section 2 covers the data, while section 3 is a simple look at the transition of corruption. Section 4 brings the A-table. Section 5 reports, and section 6 concludes.

Table 1. Two models: In the analysis x = *TI*, the corruption perception index

Model	Equation	Name
	Equations for the A-table: Using annual data, DP are the long run instruments	
(1)	$x_i = \beta_0 y_i + \alpha + u_i$	OLS estimate, the correlation between y and x
(2)	$x_i = \beta_0^{DP} y_i^{DP} + \alpha + u_i$	IV estimate, the causal effect of y to x
	Equation for the B-table: Using panel data	
(3)	$x_{it} = \beta_0 y_{it-1} + \alpha + u_{it}$	Basic model, the panel version of (1)
(4)	$x_{it} = \beta_1 y_{it-1} + \gamma x_{it-1} + \alpha_i + \alpha_t + u_{it}$	AJRY model
(5)	$\beta_0 \approx \beta_1 / (1-\gamma)$	Long-run relation between β_0 and β_1
Variables used (the β's and γ are the parameters estimated)		
i, countries	x, transition variable	y_{-1}, income lagged
t, time	x_{-1}, lagged transition variable	u, residuals
α, constant	α_i, fixed effects for countries	α_t, fixed effects for time

Note: An (i,t)-panel is needed to estimate (3) and (4). In the panel each cell should have three observations (x, x_{-1}, y_{-1}).

2. The data: An unbalanced panel of four periods and 156 countries

Transparency International is an international NGO that publishes the TI Index of perceived corruption. The index now covers the T = 15 years from 1995 to 2009.[3] We divide the 15 years into five periods of three years: 1995-97, 1998-00, 2001-03, 2004-06 and 2007-09.

The data cover M = 184 countries for T years. The number of countries covered is rising over time from only 41 in 1995 to 180 in 2009. If the data had been complete they would have reached $N = 15 \cdot 184 = 2760$ observations. Table 2 shows that N = 1750 observations are available. Thus, no less than 36.6% of the complete data are missing. Only 38 countries have all 15 observations.

For the calculation of the period averages, observations have been accepted, even if only one of the three years is covered. Once the five 3-year averages are formed, the income for the years 1997, 2000, 2003 and 2006 are included as the initial income for the periods 1998-00, 2001-03, 2004-06 and 2007-09 of the panel. The data for income is the natural logarithm to GDP per capita from the Maddison data set.

Model (4) includes the lagged TI-index, hence the panel loses the first period. Also, countries which only have observations from one period must be deleted. Some (mainly small) countries have no income observation, but we managed to get a panel with four periods and 156 countries, where all data (TI, TI_{-1}, *income*) are available for 429 cells. A full panel of 156 countries and four periods has 624 cells. Thus, the panel has 31% missing values

Table 2. All observations for the TI Index: The number of years T = 1, … ,15, for each country M = 1, … ,184, and the number of observations $N = L \cdot M$

T	*M*	*N*	*T*	*M*	*N*	*T*	*M*	*N*
1	2	2	6	12	72	11	13	143
2	2	4	7	26	182	12	20	240
3	15	45	8	6	48	13	9	117
4	11	44	9	8	72	14	9	126
5	9	45	10	4	40	15	38	570
						Sums	184	1750

3. The annual data are published towards the end of the year. The 2009 data were published November 17th 2009.

3. A look upon the transition from corruption to honesty

Figure 1 shows the 429 3-year averages for the TI-index over the initial income data. This figure has been made for various years and with various country-groups marked. It always looks very much like the graph below, which has two well-known features. First, the scaling is made so that the TI index rises with honesty, i.e. when corruption falls. Second, the shape of the point scatter always looks like a banana.

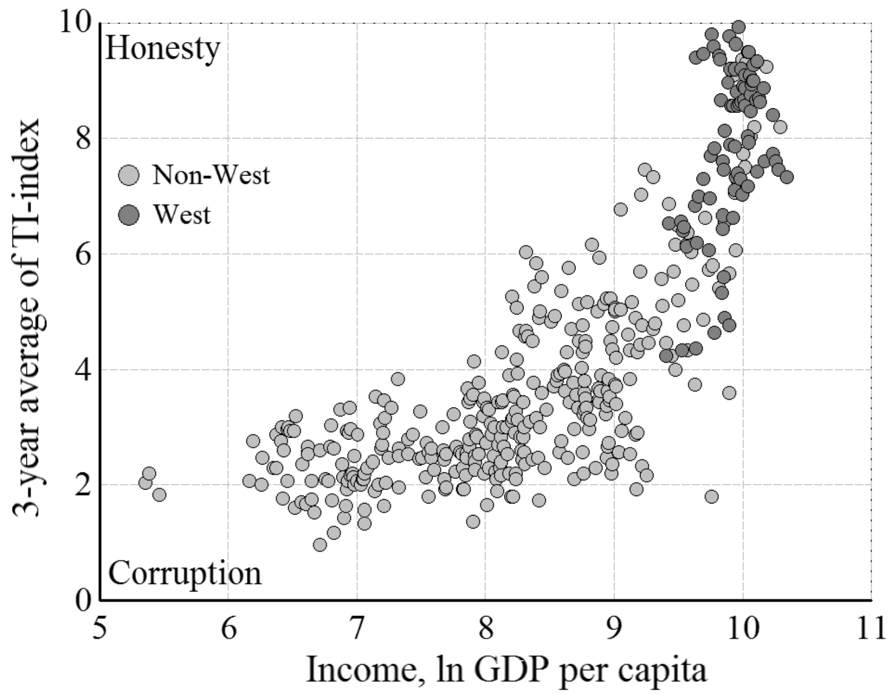
The cross-country pattern has three notable features:

(i) The slope is positive throughout The full rise of log incomes from about 6 to about 10½ points thus causes corruption to fall from a high level of about 2 index points to a low level of about 8½ index points. This is an increase of a little more than 1.4 points per point increase in log income. Hence regression coefficients β would have to have to have about this size to explain the transition from corruption to honesty.

(ii) The slope is upward bending. It is easy to correct for, but once the lagged endogenous variable is included, the deviation from linearity no longer matters.

(iii) Most of the bend can also be explained by a dummy variable for Western countries.

Figure 1. The transition of corruption



4 Table A: The long run causality test

The long-run causality test is reported in Table A. It will be interpreted in three steps:

Table 3. The A-table for transition of corruption

Time *t* is 1995	Main model	Robustness of model to instrument variation			
Dependent variable: *TI*	(1)	(2)	(3)	(4)	(5)
No. of obs. (countries)	98	103	98	98	141
			OLS estimates		
β_0 on income, y_{it},	**1.36**	**1.47**	**1.36**	**1.36**	**1.41**
	(12.8)	(14.0)	(12.8)	(12.8)	(15.3)
Centered R²					
	IV estimates: *y* is instrumented with the DP variables				
β_0^{DP} on income, y_{it},	**1.52**	**1.57**	**1.49**	**1.27**	**1.29**
	(9.3)	(10.6)	(9.2)	(8.7)	(10.0)
DP-instruments [a)]	*Biofpc, geofpc*	*Bioavg, Geoav*	*animals, Plants*	*axis, size, climate*	*coast, frost, maleco*
	Hausman test for parameter consistency of OLS and IV estimate				
C-statistic (p-value)	0.20	0.34	0.28	0.38	0.17
			Tests of validity of the IV-procedure		
First stage partial R^2	0.43	0.51	0.43	0.53	0.52
Sargan test (p-value)	0.90	*0.34*	0.23	0.11	**0.02**
	Cragg-Donald test for the strength of the instruments in the IV estimate				
Presumed causality: $y \Rightarrow \Pi$	36.09	52.50	35.86	35.17	49.60
CD critical value (size)	19.93	19.93	19.93	22.30	22.30
Reverse causality: $\Pi \Rightarrow y$	**24.05**	**31.26**	**23.35**	13.35	18.84

a) The DP-instruments are discussed and documented in Gundlach and Paldam (2009a). Estimates in bold are significant at the 5% level.

Step 1: Are the IV estimates valid? All Sargan-tests except one do not reject the exclusion restriction and all the CD-tests show that the instruments are not weak. Also, the first stage partial R^2s are satisfactory. Hence we can trust the IV-estimates.

Step 2: Is the level of income causal to the level of corruption? The IV estimates show that $\beta_0^{DP} \neq 0$. Consequently, causality is established.

Step 3: Is the estimated effect of income on corruption β_0^{DP} large enough to explain the transition seen on Figure 1? Three conditions have to be met.

- (c1) The correlation estimated by the OLS-estimate is the same as the causal effect found by the IV-estimate, i.e., $\beta_0 \approx \beta_0^{DP}$ by the Hausman C-test.
- (c2) The joint estimate of β is about 1.4, as was needed by (ii) in section 3.
- (c3) The CD-test statistics indicate the presence of strong instruments; they are substantially larger than for a counterfactual reverse transition.

From this we conclude that income fully explains the transition of corruption

5. Table B: The transition comes and goes

Table 4 is calculated as the B-tables in Gundlach and Paldam (2010). It shows the panel version of the basic model (3) as column (1), and the AJRY-model (4) as column (2). The basic model (1) shows a strong transition as in the A-table. In the AJRY model income is supplemented with the lagged endogenous and the two sets of fixed effects for time and countries. This makes the effect of income vanish. This is the very same picture that was found for the three other transition variables in Gundlach and Paldam (2010). The remaining columns, (3) to (7), show models with different combinations of the three AJRY controls.

The seven estimates of β_0 in the top line of estimates are somewhat inconsistent. The fixed effects for countries and the lagged dependent variable are strongly correlated, and when there is no fixed effects for countries in the relation the coefficient to the lagged dependent becomes precariously close to 1, so the equation has a near unit root.

Table 4. The effect of income on the TI-index.

Column (1) is the estimate of model (3) and column (2) is the estimate of model (4)

Dependent variable: *TI*	Basic model	AJRY model	Mixed model variants				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_0 on income, $y_{it\text{-}1}$	**1.622**	[0.187]	[**2.039**]	0.306	[**0.373**]	**0.433**	[**2.059**]
	(24.1)			(0.93)		(2.08)	
β_1 on income, $y_{it\text{-}1}$, adj.		0.147	**0.208**		***0.373***		**0.208**
		(0.47)	(3.54)		(1.85)		(3.16)
Lagged dependent, $TI_{it\text{-}1}$	No	**0.215**	**0.898**	No	**0.211**	No	**0.899**
(t-ratios)		(4.49)	(33.2)		(4.50)		(35.6)
Country fixed effects	No	Yes	No	Yes	Yes	Yes	No
Time fixed effects	No	Yes	No	Yes	No	No	Yes
Number of observations	429	429	429	429	429	429	429
Number of groups		150, 4		150, 4	150	150	4
R^2 within		0.108		0.034	0.097	0.025	0.950
R^2 between		0.948		0.599	0.884	0.607	0.999
R^2 overall	0.634	0.938	0.950	0.620	0.893	0.634	0.952

Note: Data are as explained in section 2. The t-ratios in the ()-parentheses are calculated from robust standard errors. Estimates in bold and italics are significance at the 10 percent level. Coefficients in []-brackets are estimates from equation (5). Estimates in bold only are significance at the 5 percent level. (1) and (3) as simple OLS-estimates, while the other estimates are panel regressions, with. The R^2s are not fully comparable for the panel OLS and the simple OLS regressions.

Table 5 The power of the three controls from the AJRY model

The column ΔR^2 gives the increase in the R^2 from including income

The three levels	The model estimated	R^2	ΔR^2	N	Variables	Df
Level 1: Only lagged endogenous	$TI_{it} = \gamma\ TI_{it\text{-}1} + \alpha + u_{it1}$	0.948	0.948	429	2	427
Level 2: Add Fixed Effects for countries	$TI_{it} = \gamma\ TI_{it\text{-}1} + \alpha_i + u_{it2}$	0.996	0.004	429	158	271
Level 3: Add Fixed Effects for periods	$TI_{it} = \gamma\ TI_{it\text{-}1} + \alpha_i + \alpha_t + u_{it3}$	0.997	0.001	429	162	267

Note: For easy comparisons these regressions are done as simple OLS.

Table 5 shows the explanatory power of the three controls, as a cumulative sequence. First the lagged endogenous variable is alone, then the 156 fixed effects to countries are added and finally the four fixed effects for periods are also added. The table shows why the three controls are so successful. They already explain all the information in the series. Hence, there is nothing left for income to do. If income manages to get a coefficient, it will surely have so much correlation with the three controls that it becomes statistically insignificant as a determinant of corruption.

5. Conclusions

The AJRY-model is well known from microeconomic studies, where it serves to reveal spurious relations. It works a bit like the Granger causality test as it explains as much as possible of the dependent variable by itself and by the panel structure itself (i.e. the two sets of fixed effects), and leaves only the "innovations" in the series. Only if the innovations in series A can explain the innovations in series B, we would accept that A causes B according to this method.

Thus, the innovation of the AJRY paper is the application of the micro causality test to the macro field of growth and development, where it appears to show that *all* long-run transitions are spurious. Growth and development is a field with much multicollinearity, where we look for long-run relations, and rather desperately try to sort through the maze of multicollinearity to find the basic patterns. The fact that such long-run relations are identified as spurious by the said test is perhaps not so surprising, but it remains an open question which insights about long-run development can be gained from it.

References:

This is a background paper to the following papers by the authors:

Gundlach, E., Paldam, M., 2009a. A farewell to critical junctures: Sorting out long-run causality of income and democracy. *European Journal of Political Economy* 25, 340-54

Gundlach, E., Paldam, M., 2009b. The transition of corruption: From poverty to honesty. *Economic Letters* 103, 146-48

Gundlach, E., Paldam, M., 2010. The agricultural, demographic and democratic transitions. Two estimation models giving reverse results. Pt working paper

It also refers to:

Acemoglu, D., Johnson, S., Robinson, J. A., Yared, P., 2008. Income and democracy. *American Economic Review* 98, 808-42

Maddison home page: http://www.ggdc.net/maddison/

Maddison, A., 2003. *The world economy: Historical statistics.* OECD, Paris. Updated versions available from Maddison homepage http://www.ggdc.net/maddison/

Paldam, M., 2002. The big pattern of corruption. Economics, culture and the seesaw dynamics. *European Journal of Political Economy* 18, 215-40

Transparency International home page: http://www.transparency.org

Note: Unpublished papers by the authors are posted on their home pages: http://www.erichgundlach.de and http://www.martin.paldam.dk