May 19, 2019

Kernels used for beauty contests of competing causal theories

Martin Paldam, Department of Economics and Business, Aarhus University,
e-mail mpaldam@econ.au.dk, home page http://www.martin.paldam.dk

1. Introduction: The two reverse kernels

We consider two variables (x_j, y_j).[1] They allow two kernel regressions in the reverse directions.

(1) $x = \mathrm{K}(y, bw)$ and (2) $y = \mathrm{K}(x, bw)$, where *bw* is the bandwidth

Formula (1) is estimated by a two-step process: (i) The data-pair are sorted by *y*. (ii) The kernel is estimated with a smoothed MA-process on the *x*'es, with a fixed *bw*, using a kernel formula. To calculate (2) the data are sorted by *x*, and the MA-process is applied to *y*'s. The resulting two kernel-curves often looks strikingly different as, e.g., Figures 1, 2 and 3. This note argues that the difference tell us something about causality.

The same data are used to calculate the two reverse kernels, so each curve 'reflects' the other. If the series have a very high correlation, the reflection becomes a mirror, so the two curves look the same, but if the correlation moderate, the mirror blurs, and the two curves come to differ. If one of the kernels gives a beautiful picture, the reverse one becomes a blurred reflection that is less pretty. Thus, the two kernels provide a neat beauty contest pointing to the causal direction between two variables.

The argument is illustrated by the 6,965 (*y*, *P*) pairs of income and the Polity index.[2] The two variables have the correlation 0.56. Competing theories claim that the correlation is due to either (1) $y \Rightarrow P$ or (2) $P \Rightarrow y$, as per kernels (1) and (2). To make the contest fair it is important to choose the *bw*'s that makes the two curves look as well as possible – this is done in section 2.

1. The data are reached from merging a panel of *M* countries into a string.
2. The data are the *Main* sample from Paldam and Gundlach (2018).

Figure 1a. The two kernels drawn over *y*

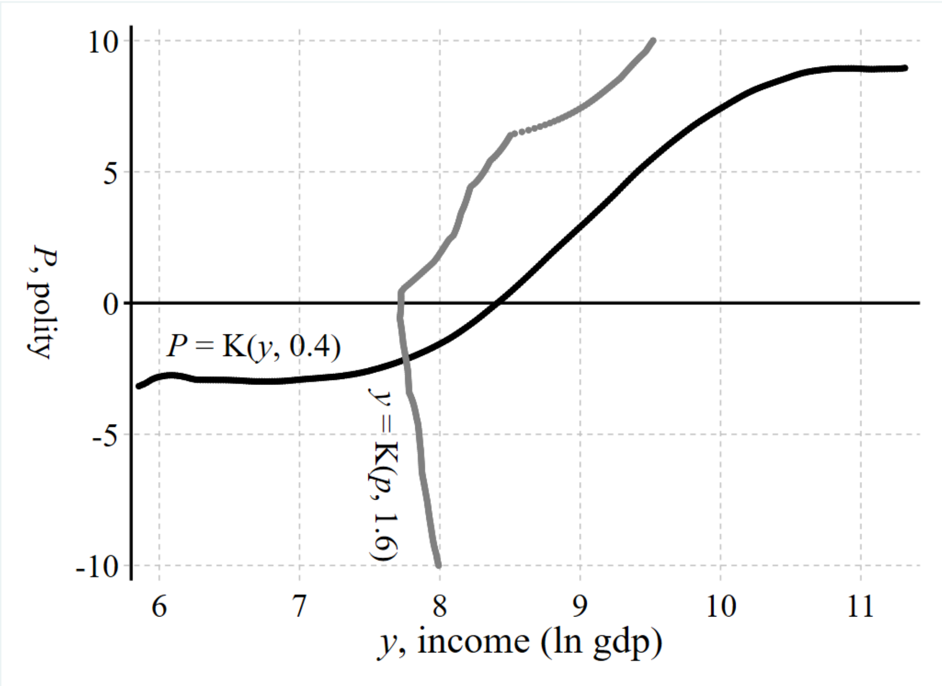
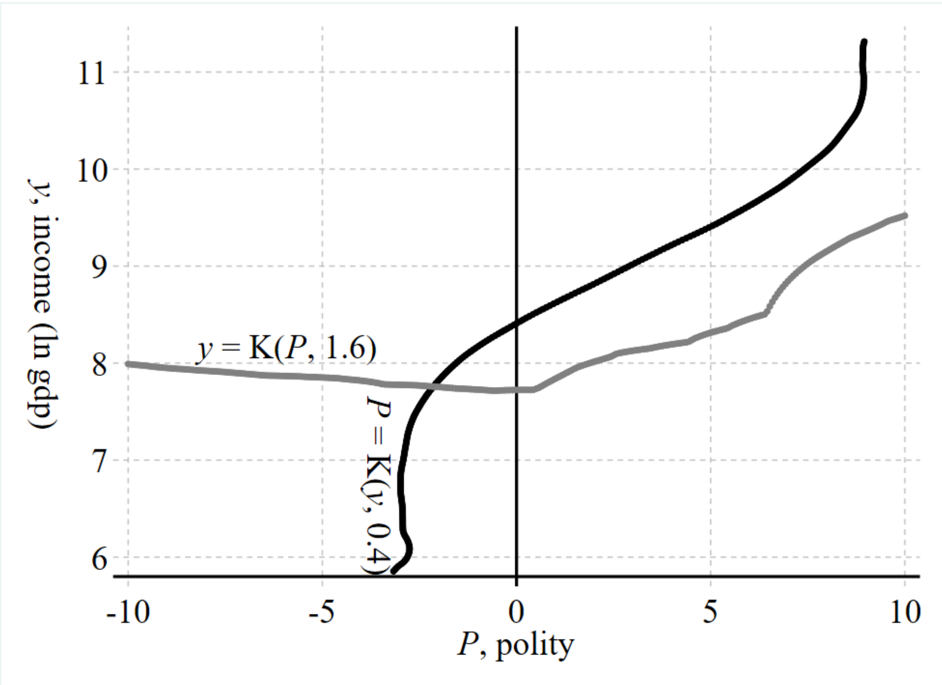


Figure 1b. The same two kernels drawn over *P*



Note: The data for the outlier Singapore are deleted. The two curves become more symmetric if the correlation of the two series gets closer to 1.

Figure 1 compare the best version of the two kernels. It show that the $P = \mathrm{K}(y, bw)$ kernel looks like a beautiful transition curve, while $y = \mathrm{K}(P, bw)$ does not look very pretty by any theory I know of. Thus, the contest has a clear winner, which is a fine piece of causal evidence.

It is important for the analysis that the process generating the string gives an effective scrambling so that successive observations from the same country are spread out, and kernels come to contain data for many countries and years. Section 4 report scrambling tests, and concluding that the scrambling is rather efficient. In principle, each $P_j = P_{it}$ may influence y_{it} to y_{it+x}, where x is a number such as 5. Perhaps there may even be repercussions from neighboring countries. So, the potential for reverse causality in the $P = K(y, bw)$ relation depends upon the weight of the observations y_{it} to y_{it+x} in $\bar{y}_j$. Section 4 show that this weight is quite small. This is a main reason why the two kernel-regressions normally looks as different as seen on Figure 1.

The rest of the paper proceeds as follows: Section 2 analyze the two kernels in more detail. Section 3 looks at the density distributions of the y and the P series, section 4 is the analysis of the efficiency of the scrambling, and finally section 5 concludes.

2 Choosing the best kernels

Figures 2a and 2b show how the two reverse kernels look when they are done with the Stata-default, that chooses an optimal *bw** by a rather arbitrary formula.

Figure 2a. The (*P*, *y*)-kernel, using lpoly from Stata with the defaults

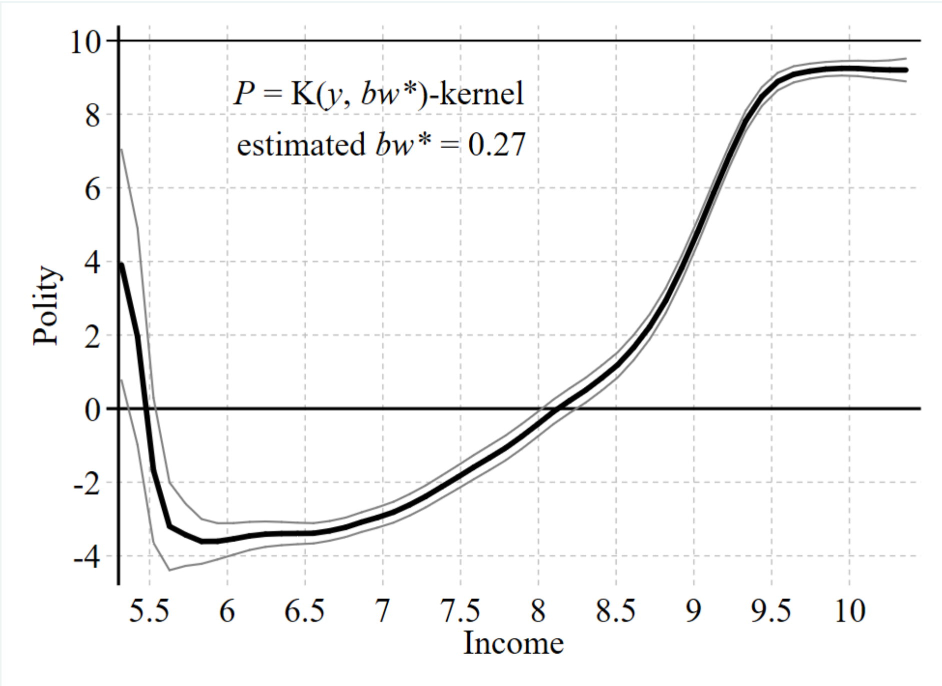
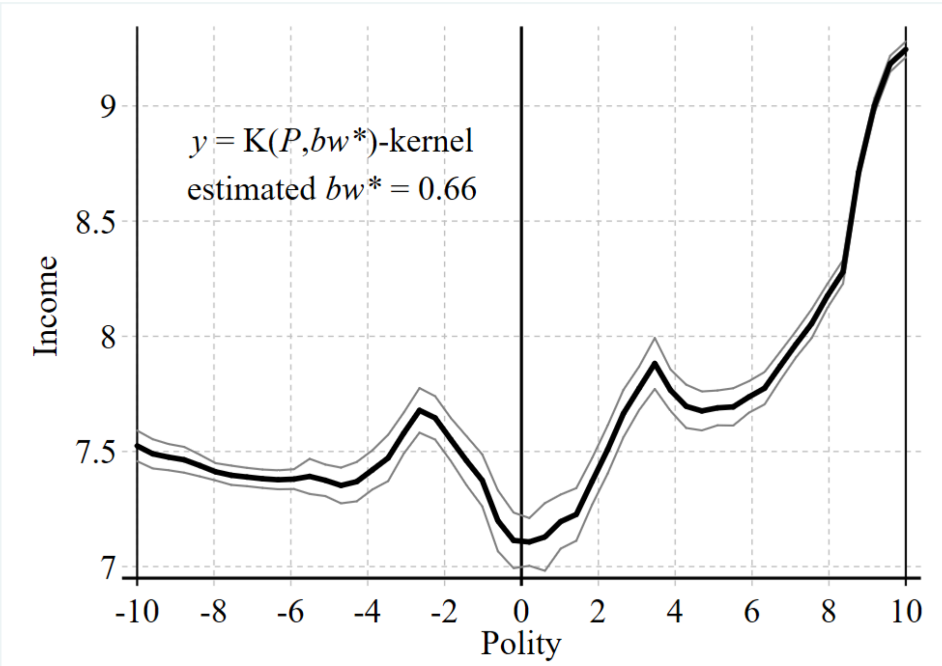


Figure 2b. The (*y*, *P*)-kernel, using lpoly from Stata with the defaults



Note: Defaults are: Epanechnikov's kernel, zero degree of polynomial smooth, *bw** is stata's choice of bandwidth. The 95% confidence intervals are included, but the scatter is deleted.

The stata formula for *bw**s chooses low values of the *bw* – just before the curve becomes wobbly. We want to choose the bw's that produce the kernels, which are the best according to theory.

The theory of the Democratic Transition predicts that *y* causes *P* and suggests that the (*P*, *y*)-kernel looks as a transition curve. The curve on Figure 2a is an almost perfect transition curve. From $y > 5.8$, the curve is as predicted by the theory, but it is problematic for $y < 5.8$. Fortunately, this part of the curve is not robust to the bandwidth as shown on Figure 3a.

The reverse theory is the Primacy of Institutions theory that also suggest a positive slope on the kernel, but it is less clear how the form should be. I have found nothing in the papers on the theory suggesting the path seen on Figure 2b. There is no trend for $P < 6$, where the curve has waves that are difficult to explain. The rise for $P > 6$ is somewhat irregular. It is consistent with the idea that the $y = K(P, bw)$-kernel is a weak reflection of the $P = K(y, bw)$-kernel, while the $P = K(y, bw)$-kernel cannot be understood as a weak reflection of the $y = K(P, bw)$-kernel.

To choose the best bw according to the theories Figure 3 shows how the two kernels looks for a broad range of *bw*'s. On Figure 3a the transition path appears to be rather robust, and the problematic part of the curve below y = 5.8 is fickle. For all bandwidths, $bw > 0.3$, the low-end problem disappears. I assess that the curve for $bw = 0.4$ is the best transition curve. It is smooth and it is flat at both low and high income. I think that it is difficult to imagine that the data could give a better confirmation of the transition theory.

Figure 3a. The $P = K(y, bw)$-kernel for a range of bandwidths

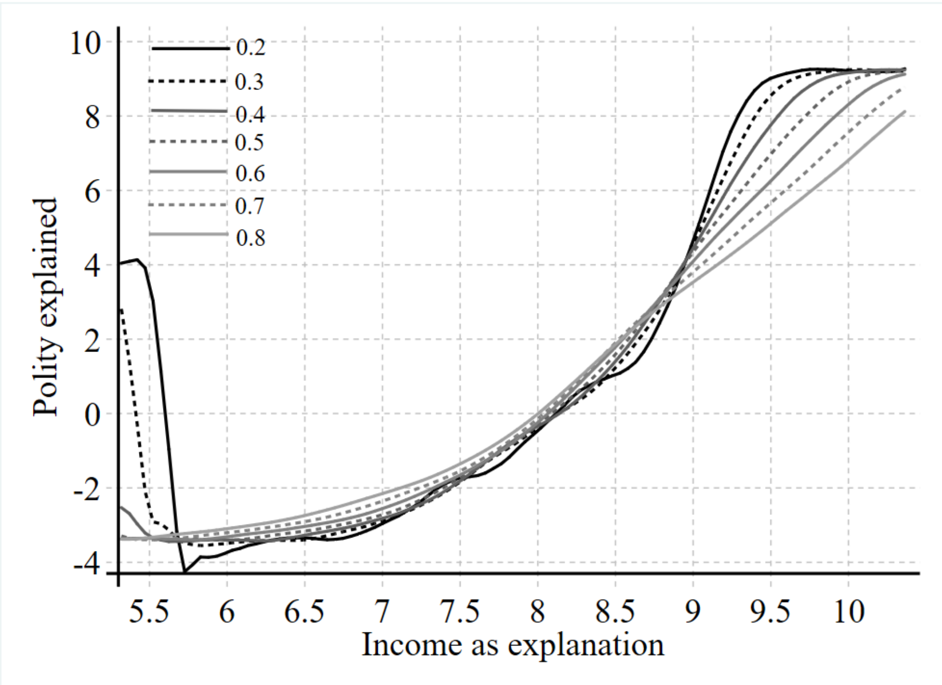
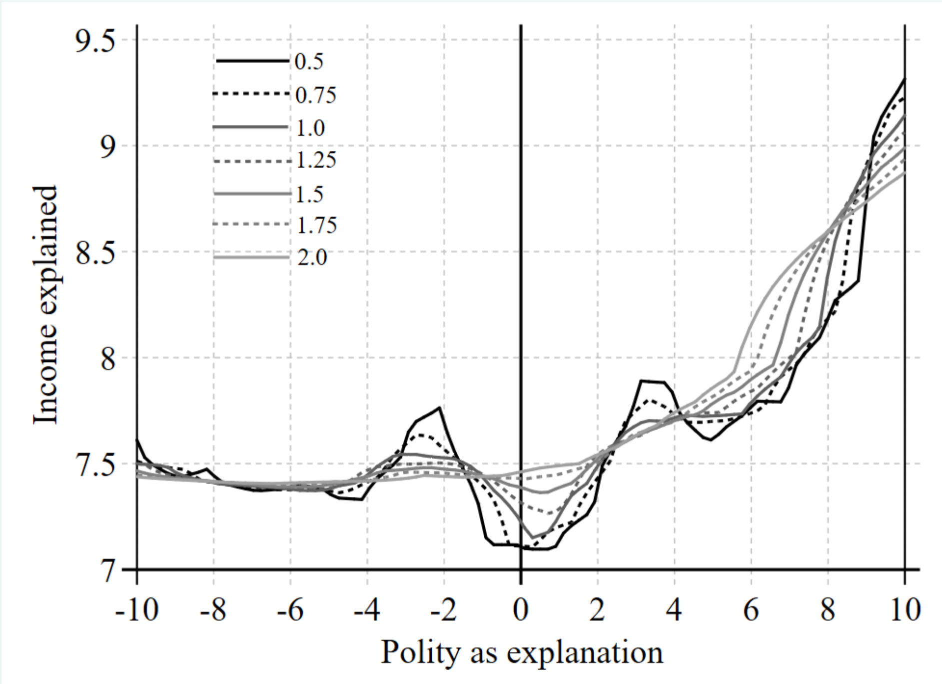


Figure 3b. The y = K(P, bw)-kernel for a range of bandwidths



Note: Confidence intervals are not included. While the range on the horizontal axis is 5 income points on Figure 3a, it 20 P-points on Figure 3b, so the range of bw's rise correspondingly.

Figure 3b presents the strange picture that is difficult to interpret. In particular, there is no robust connection from polity to income in the interval form −10 to plus 2. Once again, it is easy to interpret Figure 3b as a weak reflection of Figure 3a, but not vice versa.

Finally, I turn to the possibility that the correlation of y and P is spurious, i.e., that a third variable causes both y and P. If this is the case, it is unclear which of the two kernels should look better. One suspect that the two kernels will look much the same. The main reason why Figures 4 and 5 look so different is precisely that causality has a dominating direction. Thus, it is arguable that kernels reject that the main reason for the correlation of y and P is due to the effect of a third variable.

3. The two series: income, *y*, and Polity, *P*

Table 1 gives descriptive statistics for *y* and *P*, while Table 2 reports the reverse regressions. The high *t*-ratios on the estimated coefficients are due to the large value of *N*.

Figure 4a shown the density distribution for the *y*-variable. It is calculated by dividing the range in 20 intervals and counting the number observations in each interval. As 6,211/20 = 316, most interval holds many observations. However, for $y < 6$ and for $9.9 < y$ the data are thin.

Figure 4b reports that the Polity index has a rather strange distribution, with peaks at −7 and +10, where −7 is the score of most Communist countries before 1990, and +10 is the score of most Western democracies. The 237 observations for $P = 0$ are not part of the data analyzed in the rest of the paper, and hence displayed differently.

Table 1. Descriptive statistics for the two variables.

N = 6,211	Name	Mean	Std	Min	Max	Correlation
Income, ln to GDP per capita	*y*	7.974	1.106	5.319	10.363	0.559
Polity2 index	*P*	0.829	7.515	-10	10	

Table 2. A linear regression and the reverse

N = 6,211 Regression	(1) Constant *a*	(2) Slope *b*	(3) Reverse		(4) R^2
(1) $P = a_1 + b_1 y$	-29.4 (-51.1)	3.795 (53.1)	$1/b_2$	12.164	0.312
(2) $y = a_2 + b_2 P$	7.905 (674)	0.082 (53.1)	$1/b_1$	0.264	

Note: Parenthesis holds t-ratios.

Figure 4a. The density distribution for *y* (the natural logarithm to GDP per capita

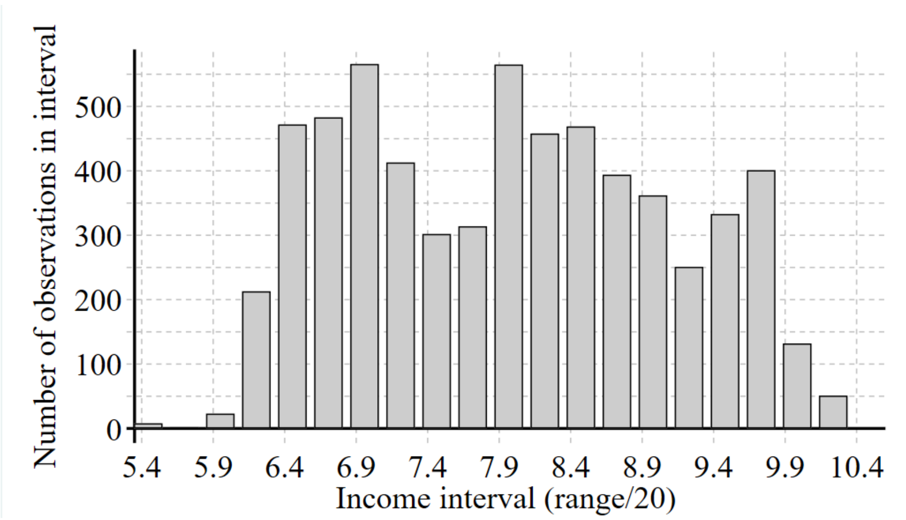


Figure 4b. The density distribution for *P*

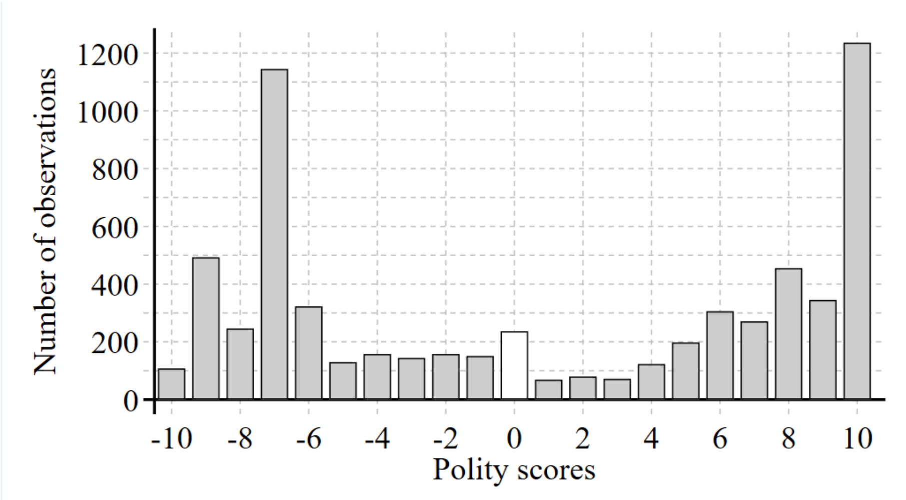


Table 3. A cross tabulation of the two variables

		Income					
Polity	**< 5.5**	**5.5-6.5**	**6.5-7.5**	**7.5-8.5**	**8.5-9.5**	**> 9.5**	**Total**
−10		2	32	67	7		108
−9		**114**	**134**	**183**	62		493
−8		18	**100**	**107**	21		246
−7		**127**	**597**	**250**	**171**		1,145
−6		45	**141**	99	32		323
−5		30	42	37	21		130
−4		26	69	49	14		158
−3		6	65	47	26		144
−2		24	44	50	19	21	158
−1		42	74	22	13		151
1	1	21	24	19	4		69
2		2	47	31			80
3	1	8	10	34	19		72
4	1	10	24	71	17		123
5	3	30	47	93	25		198
6		42	81	**102**	62	19	306
7		4	89	**117**	61		271
8		9	**116**	**149**	**153**	28	455
9		9	33	**120**	**142**	41	345
10			23	92	**511**	**610**	1,236
Total	6	569	1,792	1,739	1,380	725	6,211

Table 3 is a cross-tabulation of the two density distributions. Cells with more than 100 observations are bolded. The two disjoint bolded areas in the table are due to the two-peaked distribution of *Polity*.

4. The stacked and sorted data: How well are strings scrambled?

The introduction argue that the strings should be effectively scrambled. This is understood in two ways:

(1) The strings does not contain many sequences that are also sequences in the original data, i.e., if $y_j = y_{it}$, then it should be rare that $y_{j+1} = y_{it+1}$, or $y_{j+1} = y_{it+2}$, or … $y_{j+1} = y_{it+5}$

Table 4 count the frequencies of such sequences. They are actually rare.

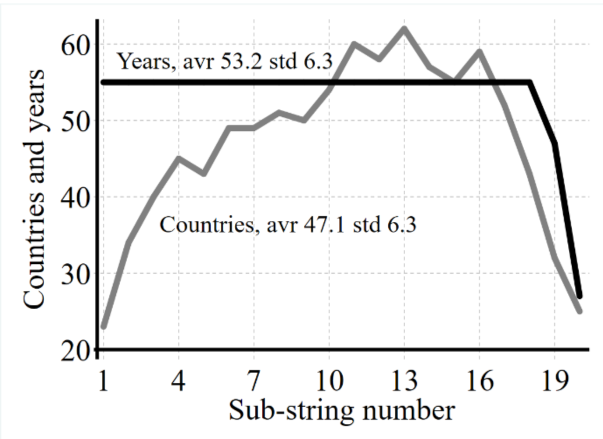
(2) The numbers of countries and years in each kernel are large.

In Figure 2a *bw* = 0.27, which is 5% of the range or 348 observations, so that the average *N* used for each kernel is 348, but it is fewer when the data are thin. So, the string is divided into 22 sub-strings of 348 observations each, where the excess 4 observations are given to the last. Then I count the number of countries and years contained in each substring. Figure 5 reports the results. There are almost 1/3 of the 155 countries and all years included in each sub-string except in the last two. Thus, both scrambling tests show that the scrambling is satisfactory.

Table 5. The scrambling tests done for string of all *N* = 6,965

Comparing *yj* and y_{j+n}	y_{j+1}	y_{j+2}	y_{j+3}	y_{j+4}	y_{j+5}	Sum
Same country	309	319	312	315	306	-
Next year	157	173	172	119	140	-
Both same	34	26	25	14	6	105
Both in %	0.49	0.37	0.36	0.20	0.09	1.51

Figure 4. The number of countries and years included in the 23 sub-strings



5. Conclusion

Thus, I conclude that that a comparison of the (P, y)-kernel and the (y, P)-kernel give evidence in favor of the Democratic Transition, and hence support for the idea that y causes P.

The evidence appears reasonably strong, but it is reassuring that it is supported by a formal causality test in Gundlach and Paldam (2009).

Reference:

Gundlach, E., Paldam, M., 2009. A Farewell to Critical Junctures. Sorting out the Long-Run Causality of Income and Democracy. *European Journal of Political Economy*. 25(3), 340-54

Paldam, M., Gundlach, E., 2018. Jumps into democracy. Integrating the short and long run in the Democratic Transition. *Kyklos* 71(3), 456-81