November 15th – 2023

(version 4)

The kernel-pair method of causal analysis of macro relations

Martin Paldam, Aarhus University, Denmark[1]

Abstract: This note looks at kernel regressions on large data sets obtained by unifying panel data. If the kernel-curve has the functional form predicted by a theory, it is a test of that theory and the causality implied. The test is strong if the form predicted is distinct and the confidence intervals are narrow. When two variables are correlated, competing theories may claim the opposite causality. The two opposite kernels often look strikingly different. When one looks as predicted by its theory, the other tends to make less sense. The first one is the obverse, while the other is the reverse. The obverse is causal evidence for the theory it reflects. Like other causality tests, the evidence is not always clear.

Keywords: Causal pairs, kernel regressions, unified data

Jel code: C18, O10

[1] Department of Economics and Business, *Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark.*

For MP: *E-mail: mpaldam@econ.au.dk. URL:* http://www.martin.paldam.dk.

I am grateful for discussions with my econometric adviser Allan Würtz.

1. Introduction

The note deals with the application of the well-known technique of kernel regressions.[2] It argues that the technique has a wide scope in *long-run macro analysis*, where it is rarely used.[3]

Research often starts from a hunch and a correlation, such as $\alpha = \text{cor}(x, y)$. It may have four causal explanations listed in Table 1. A and B are the main cases, where A is explained by the A-theory, and B by the B-theory. Thus, A and B are alternatives with the reverse causality.

In macro the unit is a country. Thus, the data for the (x, y) analysis is an (x_{it}, y_{it}) panel, with the two dimensions: i country and t time. If the theory is general and the data have the same scale, they may be unified. The panel (x_{it}, y_{it}) becomes (x_j, y_j) with $j = it$ rows. They have no natural order. A key point about kernel regressions is that they order the rows by the explanatory variable. This also means that any other variables – including time – are scrambled.

Thus, the kernel $y = K^y(x, bw)$ shows how the path of $y(x)$ looks in the average country. The connection has a ***soft*** lag structure. Macro-theories are often fuzzy. They mostly hold and the lag structure is often soft. Kernel regressions are a fine tool to study such relations.

Economic theory makes qualitative predictions. They are often vague, such as the sign of a slope, but some theories make predictions about the *functional form* of the relation. The prediction may even be so ***distinct*** as a specific non-linearity. Kernel regressions estimate the functional form of a relation, and they come with 95% confidence intervals. The estimate assumes no economic theory and barely restricts the functional form.

When a curve with the form predicted by the theory can be drawn within the 95% confidence intervals of the kernel-curve, it is a test of the theory, but it is not easy to calculate the strength of the test. The test is strengthened under two conditions:

(C1) The confidence intervals are narrow. (C2) also shows if the data unification is justified.
(C2) The prediction is distinct, preferable with clear non-linearities.

Table 1. The four possibilities explaining a significant cor(x, y)

Case		Causal structure	Interpretation	Kernel
Main	A	$x \Rightarrow y$	Model $y(x)$, the A-theory	$y = K^y(x, bw)$
	B	$y \Rightarrow x$	Model $x(y)$, the B-theory	$x = K^x(y, bw)$
Other	C	$x \Leftrightarrow y$	Simultaneity	Unclear
	D	$z \Rightarrow x$ and $z \Rightarrow y$	(x, y) relation is spurious	Unclear

The bandwidth, *bw*, is a fraction of the range of the of the explanatory variable.

[2] In November 2023 Google Scholar gave two million hits to '*kernel regression*'.
[3] The closest related paper I have found is Vinod (2017).

2. The causal pair of kernel regressions

This section considers the main case of two alternative theories explaining a correlation: A claims that the relation is $y(x)$, while B claims that it is $x(y)$. The relations are analyzed by the kernels $K^y(x, bw)$ and $K^x(y, bw)$, on a large, unified data sample with N rows (x_j, y_j). It is assumed that the confidence intervals on the two kernels are narrow, so that the unification is justified. First, we look at one of the two kernels:

2.1 The kernel regression $x = x(y) \approx K^x(y, bw)$

When the theory $x = x(y)$ is analyzed by the kernel regression $K^x(y, bw)$, the rows of the sample are sorted in the order of rising *y's*. The kernel is the smoothed moving average of the *x*'s with the fixed bandwidth *bw*.[4] It is a fraction of the range of the *y* variable. In the case studies of section 3, the fraction is between 3% and 9% of the range.

It is a consequence of the sorting by *y* that the order of the *x*'s is scrambled. For large *N*s, the scrambling is normally quite good. Data mining/overfitting is a big problem in economics[5] – it is too easy to confirm economic theories; see Paldam (2020). Kernel regressions are (almost) mining-proof.

The only variable to be fitted is *bw* the bandwidth, and the estimate changes in a predictable way, when *bw* varies. If the *bw* is too small the kernel curve becomes wobbly, so that it moves in ways that are sensitive to minor changes in the *bw*. If the *bw* is too large the curve becomes too linear and eventually horizontal at the average of the explained variable. However, there is normally a fair range of bandwidths where the kernel tells the same (interesting) story. The stata program used points to this range.

2.2 The kernel pair: $K^x(y, bw)$ and $K^y(x, bw)$

Thus, to calculate $K^y(x, bw)$ the data are sorted by *x*, where the order of *y* is scrambled. To calculate $K^x(y, bw)$ the data are sorted by *y*, where the order of *x* is scrambled. As demonstrated below this often gives very different pictures. *However,* each kernel may contain a ***reflection*** of the other kernel.

This allows a simple beauty test for causality. If one kernel – taken to be A – looks as predicted by its theory, it is the ***obverse***, and it follows that B is the ***reverse*** kernel and it is a

---

[4] The paper uses the command *lpoly* in stata, with the defaults. The *bw* chosen is a bit larger than the one estimated by the program. This makes the pictures slightly clearer.

[5] In June 2023 Google scholar has 5.2 million hits to *publication bias* and 600,000 hits to *replication crisis in the social sciences*, where app one-third is to economics.

mixture of the reflection and the B-theory. This is unlikely to be pretty. It allows us to say that the theory explaining the obverse kernel is a better explanation of the data. Hence, the main causal structure behind the correlation is revealed. The paper illustrates this idea by six cases. They confirm two methodological points:

(i) The kernel-technique is at its best analyzing large patterns in the data, notably long-run processes where it does not matter that the lag structure is soft.

(ii) The size of the reflection depends upon the correlation, α: If α is close to ± 1, the reflection is large and two kernels become symmetrical, and thus difficult to distinguish. If α is close to zero, the reflection vanishes, and the kernels become orthogonal. Thus, it may confirm/reject both theories independently. If α is moderate, the reflection is weak.

Table 2. The six case studies

	Case 1: *P and g.* Political system and growth, 1972-2018 from MP (2023c)
P	Polity2, integer from authoritarian -10 to democratic +10. *P* = 0 for no system; see Figure 1
	Source: Institute for Systemic Peace: https://www.systemicpeace.org/polityproject.html.
	Case 2: V and y. Political system and income Figure 2, 1800-2018 from MP (2023d)
V	Polyarchy, number between 0 and 1from authoritarian to democratic. See Figure 2
	Source: V-Dem project: https://v-dem.net/
	Case 3: ***F and y.*** Economic system and income from MP (2021)
F	F-index, range [0, 10] two decimals; see Figure 3
	Source: Fraser Institute: https://www.fraserinstitute.org/studies/economic-freedom.
	Case 4: *T and y.* Corruption index and income from MP (2021)
T	Corruption-index [10, 0]. It rises when corruption falls; see Figure 4.
	Source: Transparency International: https://www.transparency.org/.
	Case 5: *S* and *y.* Share of school age kids and income from MP (2023a)
S	Share of ages 6-15 years of population (%); see Figure 6
	Source: World development indicators, World Bank
	Case 6: *A and y.* Share of agriculture and income from MP (2021)
A	Share of agriculture in GDP (%); see Figure 5
	Source: World development indicators, World Bank
	The explanatory variables: Income, *y*, and growth, *g*
	GDP, is *Gross Domestic Product*, in fixed PPP (purchasing power parity) prices.
gdp	GDP per capita. The *cgdppc* series
g	Growth rate (%) for *gdp;* see Figure 1
y	*Income*, the natural logarithm to *gdp;* see Figures 2 to 6. A lp-point is a *gdp* change of 2.7 times.
	Source: Maddison Project: https://www.rug.nl/ggdc/historicaldevelopment/maddison/.

Data are annual observations. Present and former OPEC countries and Bahrain and Oman are excluded, which leaves app 137 countries.

2.3 The six cases

Table 2 lists six cases analyzed in section 3. The cases are meant to develop some intuition. The six variable-pairs are: (P, g), (V, y), (F, y), (T, y), (S, y), and (A, y). Each case is analyzed in one section by one figure with three graphs, a, b, and c.

Graph **a** has g (growth) or y (income) on the horizontal axis. Here the kernel is black – it may be explained by transition theory see section 2.4.

Graph **b** has the axis reversed, and here the kernel is gray -- here the theory varies from graph to graph see section 2.5.

Graph **c** shows both kernels together using the axes of the a-graph. Thus, the b-kernel is mirrored on the 45% line. As both graphs appear on the c-graph, the range on each axis covers the largest range on graphs a and b. If the a or b graph has a small range on one vertical axis, it looks different on the c graph.

The six examples are all chosen from prior papers of the author (MP). Thus, they are discussed at some length elsewhere, so the discussion of the cases is brief at present.

2.4 Transition theory: the skeleton of development

Kernel regression is eminently suitable for the study of transitions. They are general, underlying long-run processes overlaid with a great deal of fuzziness.

From Maddison (2001, 2003) and Galor (2011) we know that development has two basic steady states.[6] The traditional and the modern. The traditional had slow technical progress, giving growth of -10 to 20% per century. The modern steady state has convergence to the same international technical level with a stable moderate growth rate of about 2% annually.

The ***grand transition*** is the process where countries diverge from the traditional steady state converge to the modern steady state. This process typically takes a century or more.

The grand transition consists of confluent transitions in most (all?) socio-economic variables. Transitions are a long-run process, which gives the paths of the variable as a function of income. The path has a distinct non-linear form that is clearest, when the variable is a ratio. It looks as either  or , depending on the scaling of the variable. The first difference to a transition curve is hump-shaped  or , with the horizontal sections the steady states at the start and the end.

Transition theory claims that the kernel-curve is *equivalent* in two dimensions: In wide cross-country samples, and in long time series. In the cases examined by the author equivalence holds rather well. Transitions are fuzzy processes, but as large cross-country data sets for two

[6] A recent book by the author (MP 2021) deal with the grand pattern of development, ant the transitions of institutions. Hence, the present is a sketch of the core of a much larger discussion.

to five decades often are available it may be possible to make large, unified data sets. Here the kernel technique frequently generates fine transition curves, such as most of the six a-graphs below.

If the kernel for $z = z(y)$ looks like a transition curve and the confidence intervals are narrow it means that the main causal direction is from income, y, to the variable, z. If z is an index for a certain type of institution, it means that the institution is endogenous. One reason for the fuzziness of transitions is that countries make decisions on institutions and that institutions tend to stick for some time, but clear transition curves means that endogenous forces dominate in the long run.

2.5 The P-o-I alternative to transition theory

The alternative theory is consequently that the institutions generate development, i.e., the primacy-of-institutions theory. In all the first five cases, many researchers will support this alternative. However, it seems that the regressions supporting the P-o-I theory are linear, so the prediction is mainly that the slope is positive.

Transition theory is in levels: Income explains the transition variable. However, P-o-I theory explains growth, g, which is a first difference variable. Recall that $\Delta y \approx g$ is a fine approximation. It also means that $y = y_{-1} + g = y_{-2} + g + g_{-1} = \Sigma\, g$, the income level is the sum of past growth. However, the annual relation $y = y(g)$ is weak, while $g = g(y)$ is stronger. The $g = g(y)$ relation is the hump shaped transition in the growth rate, also known as the convergence equation. Paldam (2023c) analyzes the (y, g, X)-nexus, where X is a democracy index.

In cases 1 and 2 the idea is that as democracy is the best political institution. Thus, it gives faster development. In case 3 many agree that more economic freedom gives a faster development. In the same way it is often argued that less corruption causes faster development, and finally in case 5 there is an almost universal belief that more education gives faster development in due time. Only in case 6 there seems to be nobody, who argues that a reduction in the share of agriculture gives faster development. Thus, in five of the cases the alternative is clear.

2.6 The big question

When the variables are aggregates, they aggregate indicators, which often have a complex causal pattern. Still, one direction of causality may be the main one. Thus, the question examined is: ***Does the pair of reverse kernel regressions help identifying the main causal direction between two macro variables?***

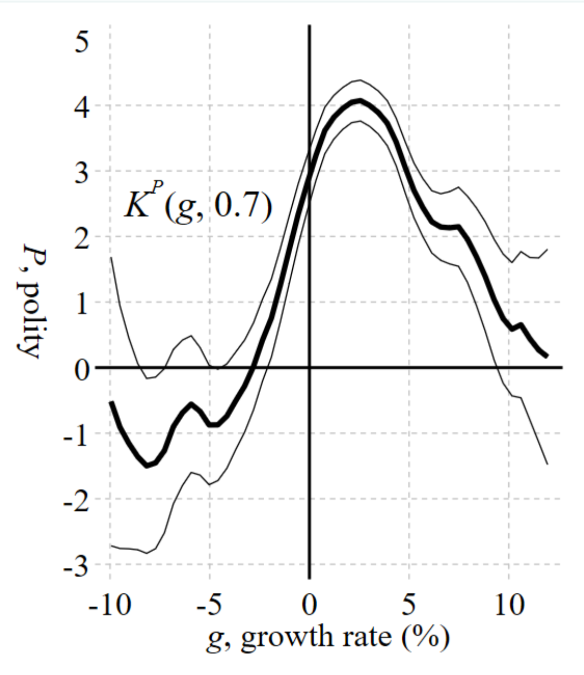
3. Six cases cross-country models

3.1 Case 1 (P, g). Polity and growth. Cor = 0.09

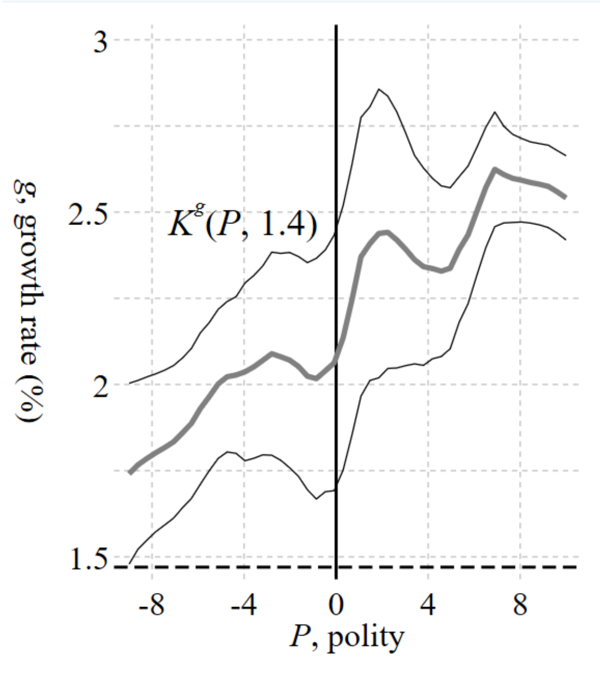
The correlation is small, so the two kernel-curves K^P (g, 0.7) and K^g(P, 1.4) are almost orthogonal. **Graph a** has the hump shape expected by a first difference transition. **Graph b** analyzes if democracy causes growth. It has a positive slope, but it has a small effect, and the kernel has wide confidence intervals. Thus, graph b may be a spurious effect of the transitions in P and g. On **graph c** it is clear that the P(g) relation is the obverse.

Figure 1. Kernel pairs for (P, g) cor = 0.08, N = 5,557

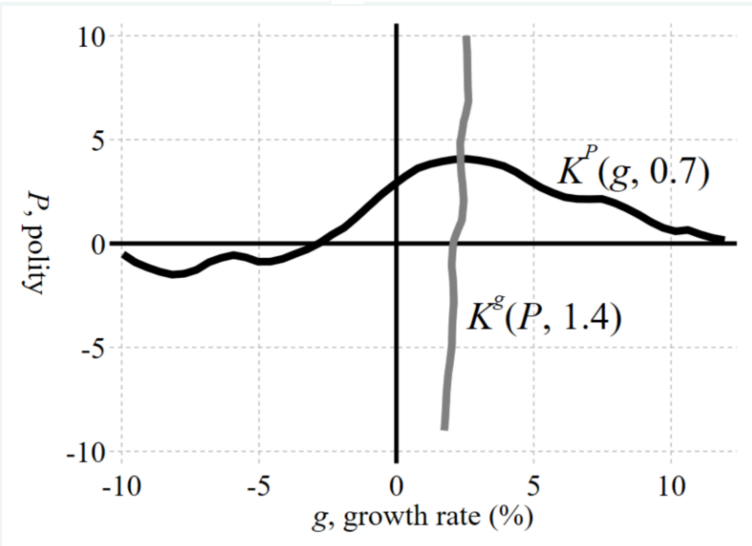
Graph 1a. $P = P(g) \approx K^P(g, 0.7)$



Graph 1b. $g = g(P) \approx K^g(P, 1.4)$



Graph 1c. Both kernels use the axes of graph 1a



Growth rates below -10 and above 12 are truncated. On Figure 1a it deletes the extreme parts of the curves. On Figure 1b it reduces the confidence intervals by about 25%, but they are still wide.

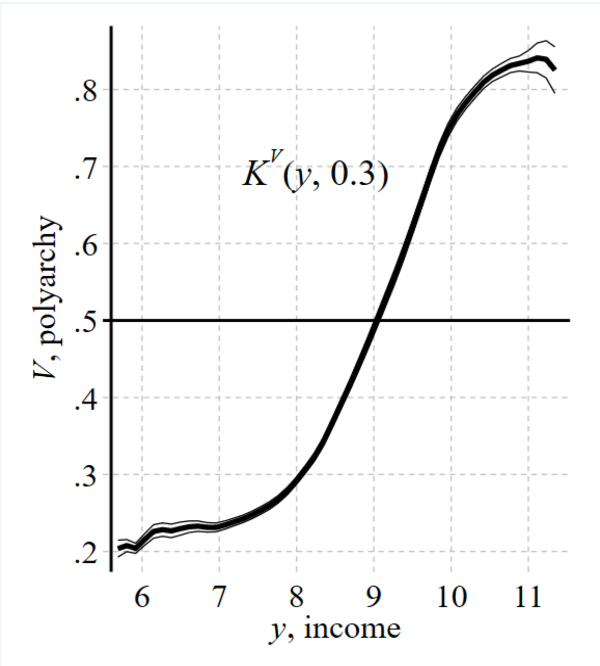
3.2 Case 2 (V, y). Polyarchy and income. Cor = 0.65

As the correlation is moderately large, the two curves are only orthogonal for $y \approx 8$, but the rise in the $K^V(y, 0.3)$-curve at the end is reflected in the $K^V(V, 0.075)$ curve. Note the narrow confidence intervals – they are particularly narrow as $N = 11{,}120$.

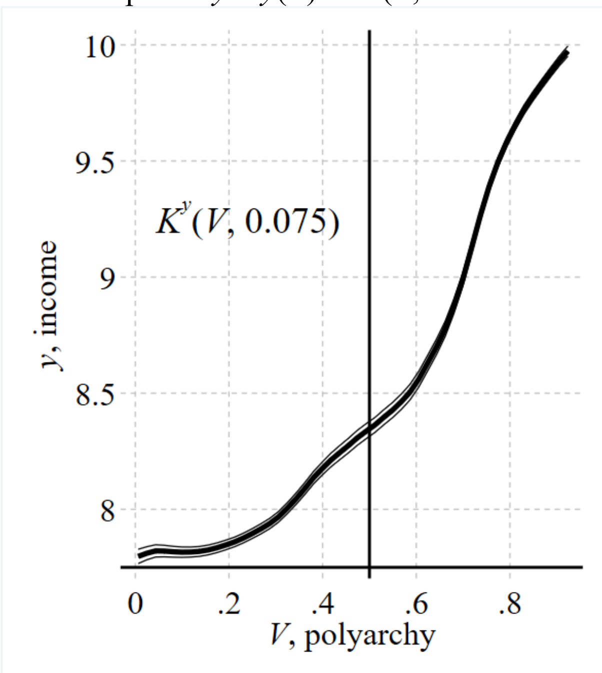
Graph a is a perfect transition curve – it is the democratic transition.[7] **Graph b** covers only the interval 7.8 to 10 on the income axis, and half of that is for V above +6, so the $y(P)$-graph does not explain much. Thus, it is obvious that the $V(y)$ relation is the obverse. The same picture appears for the polity data used in case 1 and for shorter periods as well.

Figure 2. Kernel pairs for (V, y), cor = 0.65, $N = 11{,}120$

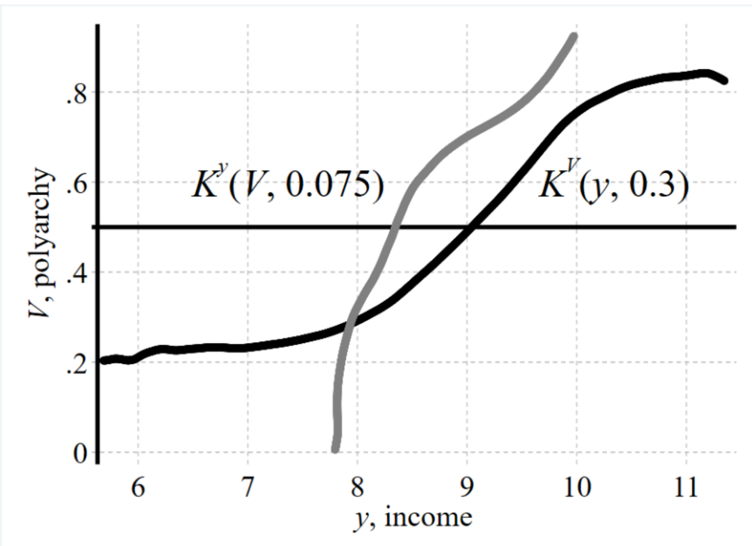
Graph 2a. $V = V(y) \approx K^V(y, 0.3)$



Graph 2b. $y = y(V) \approx K^y(V, 0.075$



Graph 2c. Both kernels use the axes of graph 2a



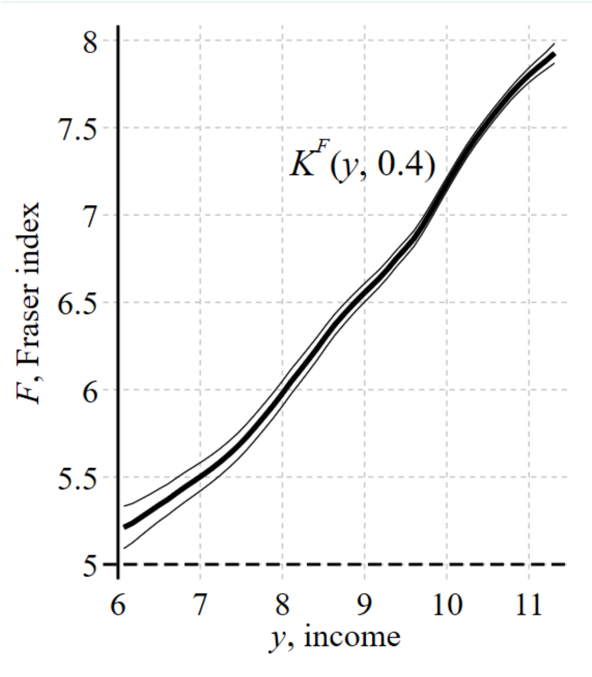
[7] Graph 2a is analyzed in Paldam (2021, 2023b and c). It is a very robust curve.

3.3 *Case 3 (F, y). Economic freedom and income. Cor = 0.69*

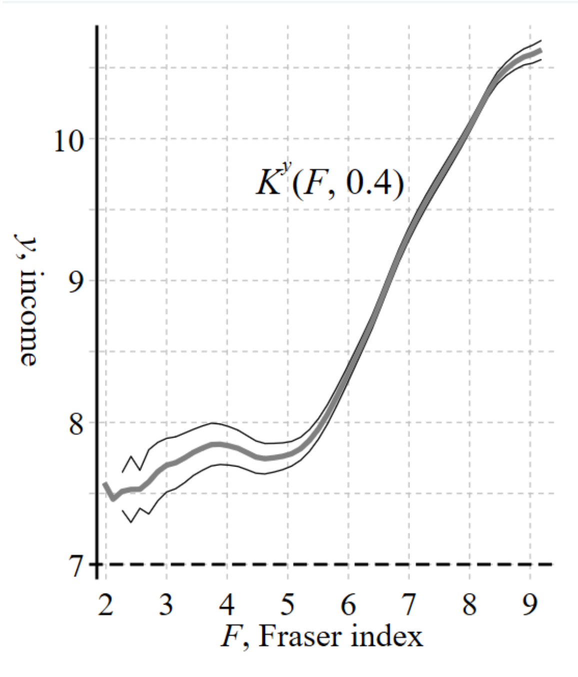
This case is less clear. The two kernels are not orthogonal at all. **Graph a** is an almost linear transition curve. **Graph b** is the income effect on economic freedom. It has a strange kink at the F-score of 5. Thus, graph a is the obverse, but this conclusion is not strong.

Figure 3. Kernel pairs for (F, y), cor = 0.69, N = 2,623

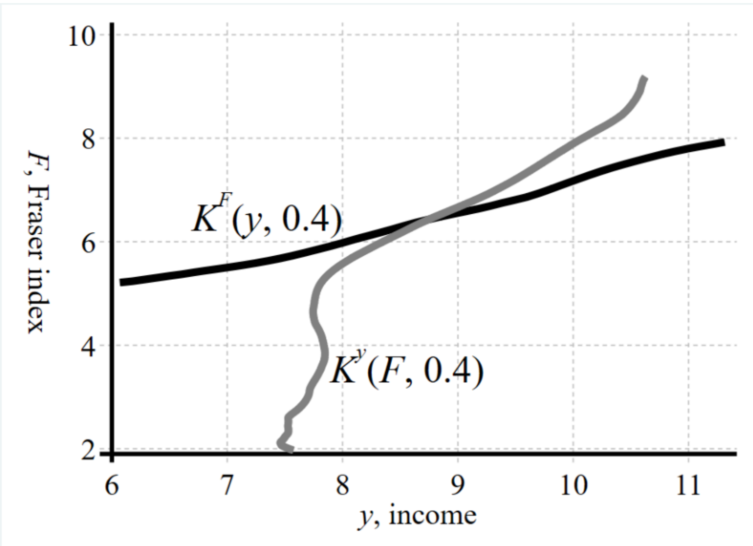
Graph 3a. $F = F(y) \approx K^F(y, 0.4)$



Graph 3b. $y = y(F) \approx K^y(F, 0.4)$



Graph 3c. Both kernels use the axes of graph 3a

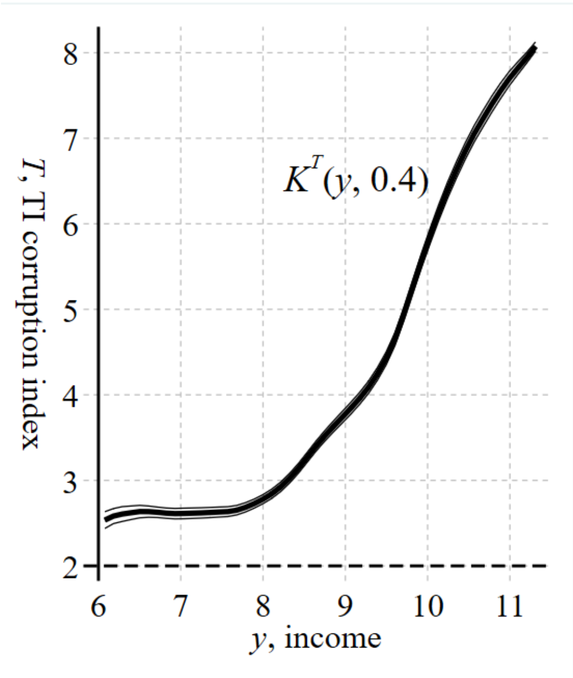


3.4 Case 4 (T, y). Corruption and income. Cor= 0.69

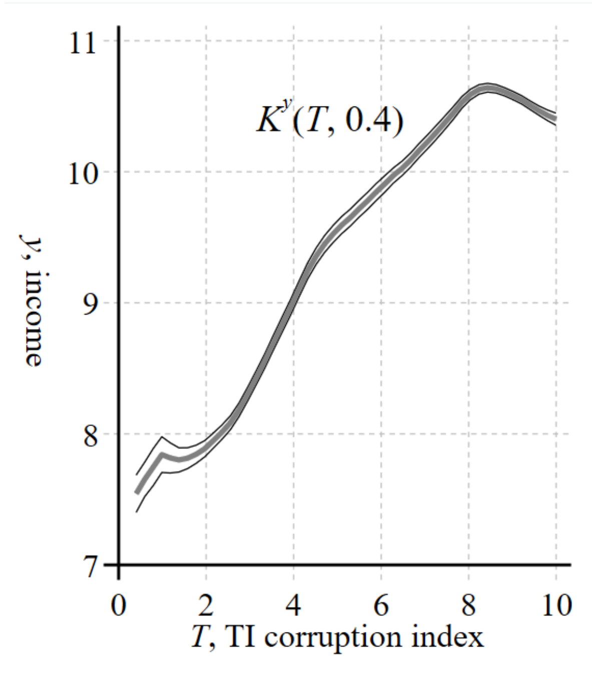
Graph a shows a nice transition, but it is incomplete. However, the index cannot exceed 10 (for no corruption), so the curve must even out for higher income. Thus, corruption has a late transition; see Paldam (2021, Cpt 10). The curve on **Graph b** has several strange bends. The a-graph is the obverse, and b is the reverse.

Figure 4. Kernel pairs for (T, y), cor = 0.77, N = 2,730

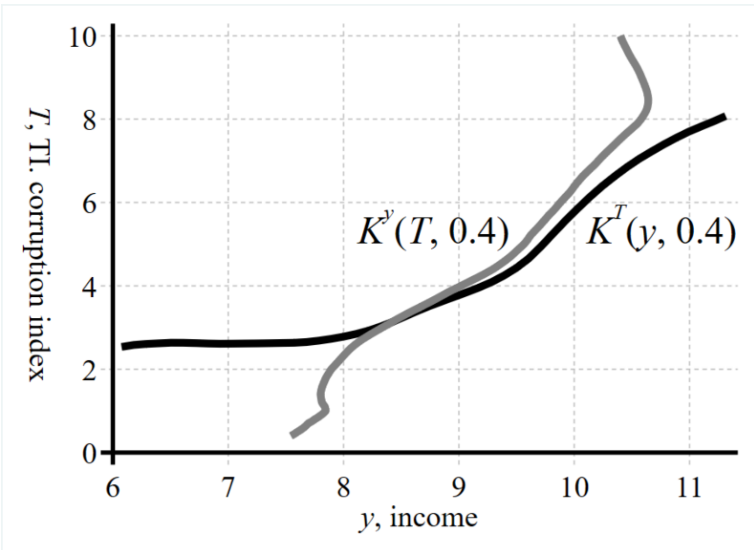
Graph 4a. $T = T(y) \approx K^T(y, 0{,}4)$



Graph 4b. $y = y(T) \approx K^y(T, 0.4)$



Graph 4c. Both kernels use the axes of graph 4a

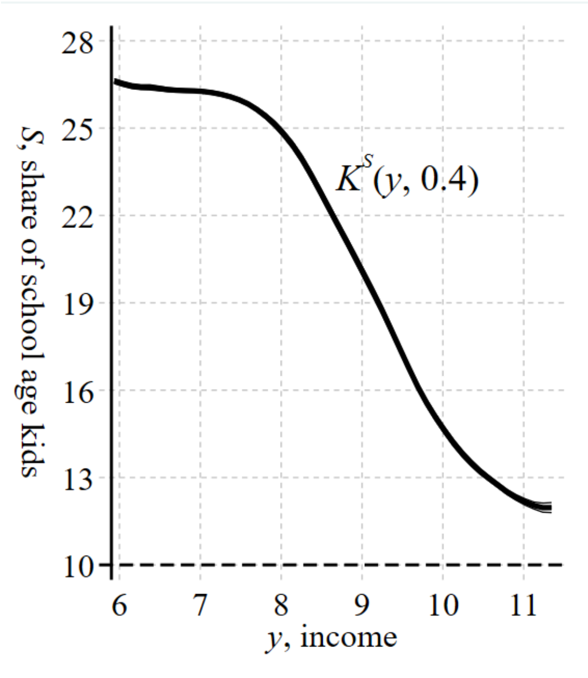


3.6 Case 5 (S, y). Share of school age kids and income. Cor = -0.83

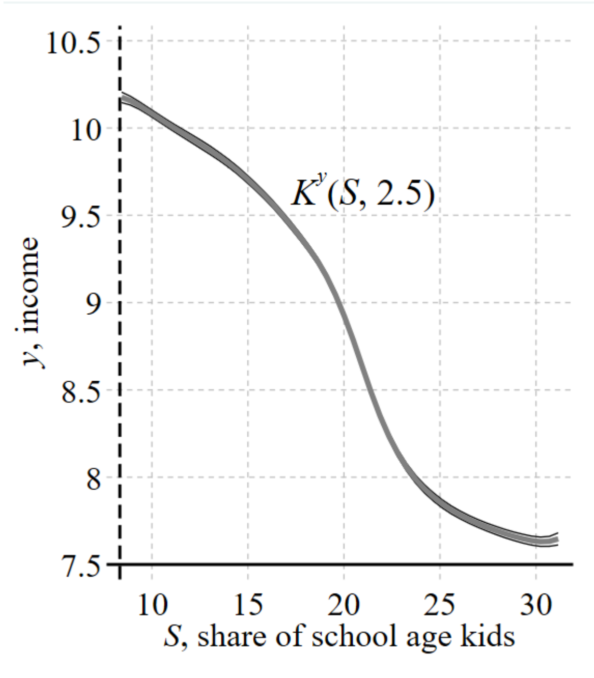
Here the correlation is -0.83, so the two graphs are similar. **Graph a** is one of the effects of the well-known demographic transition. It is easy to explain. **Graph b** is the long-run effect of the birth rate on income. It is well known that the effect is dubious. In the short run, less children increase the labor supply, and reduce the denominator in the per capita calculations of the growth rate, but in the longer run it gives problems. Thus, it is concluded that a is the obverse kernel-curve.

Figure 5. Kernel pairs for (S, y), cor = -0.83, N = 6,866

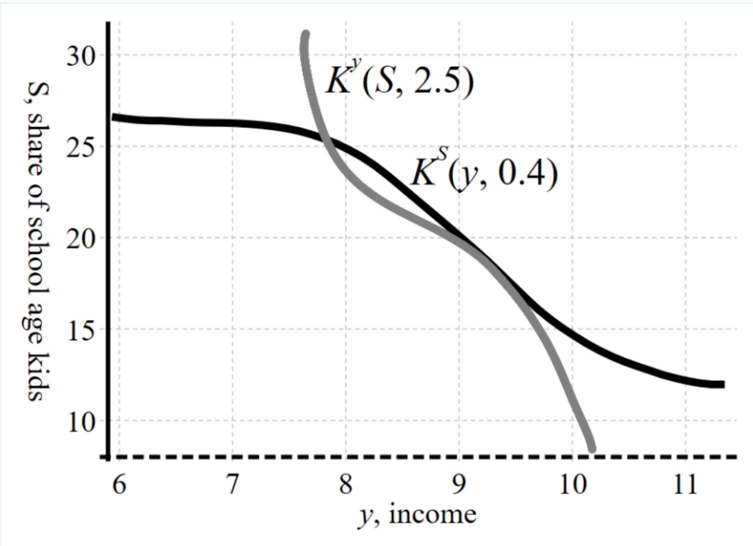
Graph 5a. $S = S(y) \approx K^S(y, 0.4)$



Graph 5b. $y = y(S) \approx K^y(S, 2.5)$



Graph 5c. Both kernels use the axes of graph 6a

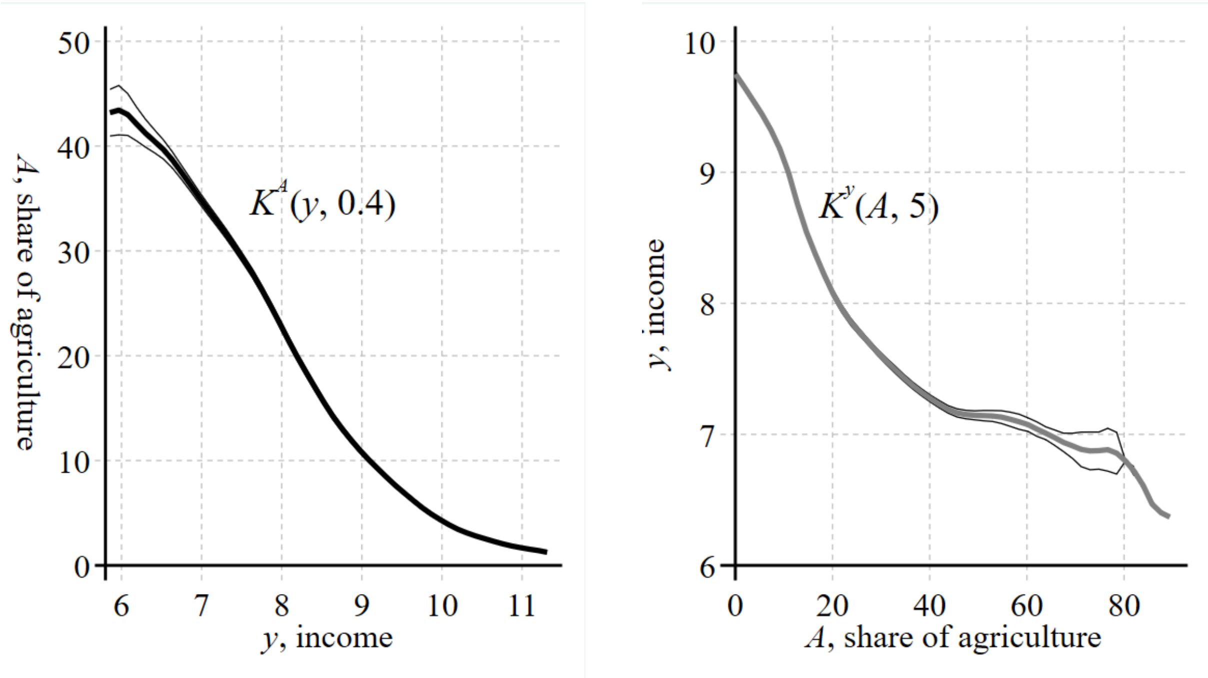


3.5 Case 6 (A, y). Share of agriculture (in gdp) and income. Cor = -0.81

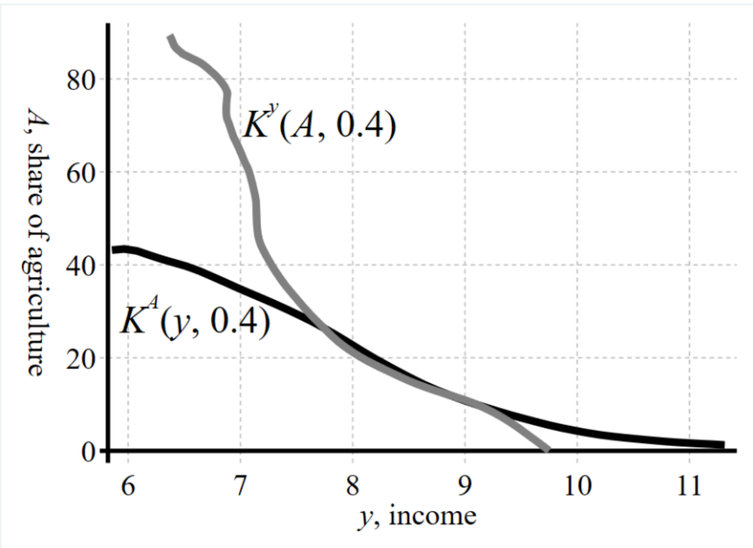
The last case is different in the sense that there seems to be a general belief that the fall in the share of agriculture is caused by development. Nobody argues that development is caused by a fall in the share of agriculture.[8] Still with a correlation of -0.81, the two graphs are similar. **Graph a** is the agricultural transition, which is well understood. **Graph b** is difficult to explain. Hence, graph a is the obverse and b is the reverse. But this conclusion is based on theory and not strong empirically.

Figure 6. Kernel pairs for (A, y), cor = -0.81, N = 5,811

Graph 6a. $A = A(y) \approx K^A(y, 0.4)$ Graph 6b. $y = y(A) \approx K^y(A, 5)$



Graph 6c. Both kernels use the axes of graph 5a



[8] There is a case for taxing agriculture to develop the modern sector, but to obtain a large tax revenue it is important that agricultural production thrive.

4 Conclusions on the economics of the six cases

In all six cases it is argued that the transition case on graph a is the obverse. In three of the cases, 1, 2 and 4, this is clear. In case 3 the transition curve is too linear, and in cases 5 and 6 the correlation is so high that the a and b graphs are similar. However, the theoretical argument for the reverse curve is weak in the two cases.

Transition theory (A) is an ideal testing ground for the kernel-pair method, as transition theory predicts that wide data sets and long time series contain the same underlying pattern with a distinct form. However, transitions are fuzzy processes. Thus, in the short-run data for few countries the fuzziness often dominates.

The alternative theories (B) differ, so they mainly appear as an alternative, but if they had been the obverse in a few cases, it would have been important. The most intriguing figure is Figure 1, where the correlation is so low that both the (A) and the (B) theory may be true.

Section 2 ended by asking if the kernel-pair method was a useful method to obtain causal knowledge about the relation between two variables. The examples show that it is! It is not a method that works in all cases, but sometimes it does, and solid causal information is always hard to come by.

References to project:

Paldam, M., 2018. A model of the rational economist, as researcher and policy advisor. *European Journal of Political Economy* 54, 6-15

Paldam, M., 2021. Methods used in economic research. An empirical study of trends and levels *Economics* 15, 28-42

Paldam, M., 2021. *The grand of development and the transition of institutions*. Cambridge University Press, New York, and Cambridge (UK)

Paldam, M., 2023a. The transition of education. A cross-country macro analysis. *European Journal of Political Economy*, online February 2023

Paldam, M., 2023b. Explaining the path of the democratic transition. The inevitable collapse of the three pillars model. WP, November

Paldam, M., 2024. Income, Growth, and Democracy. Looking for the main causal direction in the nexus. WP, January

Note on references:
Cases 1 to 5 are discussed in MP (2021), while case 6 is discussed in MP (2023a). Cases 1 and 2 are also discussed in MP (2023b and c).

Other references:

Galor, O., 2011. *Unified Growth Theory*. Princeton UP., Princeton

Maddison, A., 2001. *The World Economy: A Millennial Perspective*. OECD, Paris

Maddison, A., 2003. *The world economy: Historical statistics*. Paris: OECD. Updates on the Maddison home page see Maddison Project

Vinod, H.D., 2017. Generalized correlation and kernel causality with applications in development economics. *Communications in statistics – simulations and computations* 46(6<) 4513-34. https://doi.org/10.1080/03610918.2015.1122048