2. Some Technical Points

The nine sections of Chapter 2 introduce the techniques used in the book. First, the *five properties* of an ideal transition curve are listed (s1). A factor analysis of the four main variables gives one and only one strong common factor – it is the joint transition.

The literature shows that the usual regression tools often fail to find transitions. It is argued that the tools are wrong for the problem (s3). A more appropriate tool is *kernel regressions* on unified data samples using income as the organizing variable. The logic of this tool is explained (s4), and it is demonstrated that kernel regressions effectively scramble the panel structure, so that only the income variable matters for the form of the kernel (s5). The kernels found have most or all of the five properties within narrow confidence intervals. Thus, kernel curves are fine estimates of the transition curves. This allows (s6) the interpretation of the curves as *equilibrium* paths, so that they are attractors for the changes in the indices.

Three causality tests are used. (s7) The beauty test compares two reverse kernels – one explaining the index by income and the other explaining income by the index. It is often easy to see which explanation is closest to its theory. Correlograms for income and each index are used to show if either variable can predict the other. The DP-test is a formal TSIV-test (two-stage instrument variable) using the development potential of countries as instruments for income (s8). Finally, it is discussed if transition curves may be artifacts (s9).

Table 1. Terminology used in Chapter 2

Transition terminology for the institutional index X (a)	
Traditional steady state. All countries in 1750 and low-income countries (LICs) until recently.	
Modern steady state. High-income countries today (HICs), with the OPEC exception.	
Grand transition. The path that connects a low-level divergence and a high-level convergence.	
X_{it}	*Panel* representation of the variable, where i is the country and t is time (a).
X_j	*Unified* representation, where j is the order of the data. Divided in *Main* and *OPEC* sample.
$\Pi^X(y_j)$	*Transition* curve. X_j is sorted by y_j. Gives the net change in X, necessary for the transition.
λ^X	*Slope* of transition curve, $\lambda^X = \partial\Pi^X/\partial y$. It has the same sign, either $\leq$ or $\geq$, for full y-range.
$K^X(y_j, bw)$	*Kernel* estimate of $\Pi^X(y_j)$, with bandwidth bw. Thus, $K^X(y_j, bw) \approx \Pi^X(y_j)$.
Θ	*Tension variable,* $\Theta = X(y) - \Pi^X(y)$. If $\Theta > 0$, the country has too much X, and vice versa.
G^X*-ratio*	*Excess* movements in X. The gross movements in X relative to the net change.

Note (a): In this chapter, X is one of the bounded indices for the level of institutions P, F or T; see Table 1.2. These four variables are all scaled to increase with y like in graph (B) in Figure 1.

2.1 Variants of the transition curve, $\Pi^X(y_j)$

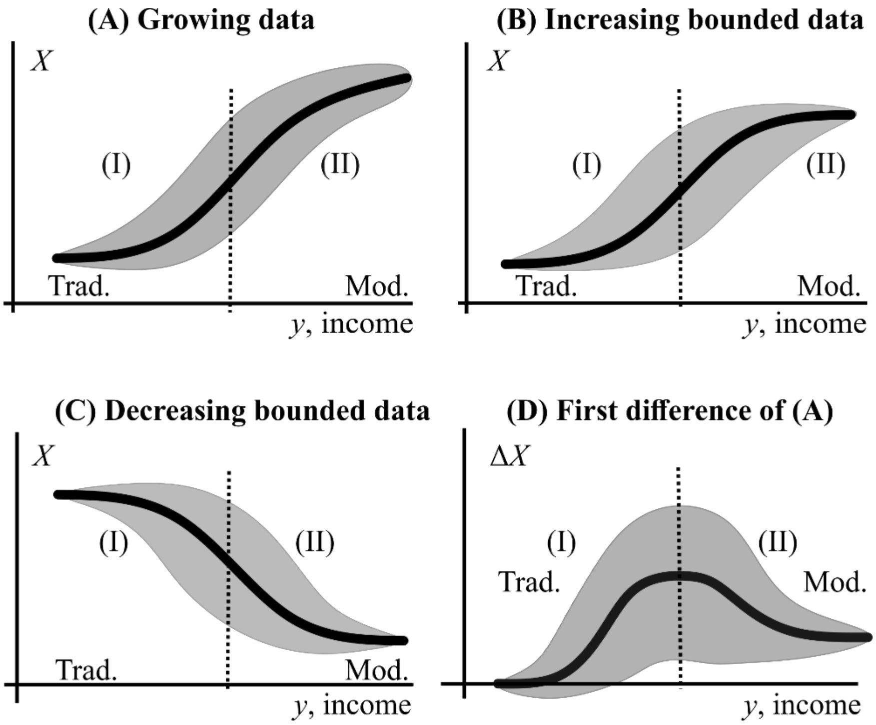
The variable of interest is termed X. Table 2 gives the characteristics of a Π^X-curve – if it has all these characteristics, it is termed ***beautiful***. Figure 1 shows how such curves should look.

Graph (A) considers a growing variable, such as a GDP component. It is constant at the traditional level, but it grows faster than the GDP during the transitions, and it ends by growing parallel with GDP in the modern steady state. The transition for bounded indices are flat at the two ends, so they look as graph (B) or (C), depending on the scaling of the data.

Table 2. Six characteristics of a beautiful transition curve for the variable X

No	Characteristic of the transition path for a variable with a bounded range.
1	Traditional (low-end) steady state level, X^T. Countries diverge from this level.
2	Modern (high-end) steady state level, X^M. Converge to this level.
3	X^T and X^M are rather different.
4	Average path is smooth between the two levels – the slope has the same sign.
5	Explanatory power is substantial, but the transition does not explain everything.
6	The transition is a causal relation from y to X, though it may have a simultaneity bias.

Figure 1. Transition curves, $X = \Pi(y)$



'Trad.' and 'Mod.' are short for the traditional steady state and the modern steady state. The vertical dotted line is where the transition is fastest. The area (I) to the left of the line has divergence, and the area (II) to the right of the line has convergence. Graph (D) is graphs (C) and (D) from Figure 1.2. The curves on figure (A) to (C) are for ***level*** variables, while (D) is for a ***first difference*** variable. The gray areas point to the fuzziness around the transition curve. It is particularly large on graph (D) for the first difference data.

As the slope of the transition curve for a *level* variable has the same sign (either ≥ or ≤), the changes along the curve are net changes adding up to the transition, where the G^X-ratio is one. In practice, most countries go through more changes, for any X-variable, giving a G^X-ratio that is well above one, as discussed in Chapter 1.

Chapter 12 analyzes the growth rate of GDP. It looks as graph (D) on Figure 1. The 'corresponding' graph in (log) levels (A) makes no sense, as it has the same variable at both axes. This reflects that the GDP is the aggregate of all sectors, so that when some increase (relatively), others have to fall (relatively). In this case, graph (D) is the interesting one. It is ***hump-shaped***, as it is the first difference of many graphs looking as (B) and (D). However, by going to the first difference, the variation around the curve becomes larger.

2.2 The Grand transition as the common factor in the four main datasets

The four main annual datasets used in the book are: (i) *y*, income, (ii) *P*, the Polity index, (iii) *F*, the Fraser index, and (iv) *T*, the Transparency corruption index. The four variables have 1,965 overlapping annual observations for the Main sample. Table 3 reports a factor analysis of the four variables. Factor1 has an eigenvalue of 2.4, while higher factors have eigenvalues far below the acceptable level that is normally set at 1. Thus, the analysis shows that these four variables have ***one and only one common factor***. It loads strongly to all four variables. I claim that it is the ***Grand Transition***. The claim will be supported as the book proceeds.

Table 3. A factor analysis of the four annual variables *T*, *y*, *F* and *P*

Importance of factors			Factor loadings		
Factor	Eigenvalue	Cumulative	Variable	Factor1	Factor2
Factor1	2.416	1.098	*T*-index	0.856	-0.048
Factor2	0.018	1.106	*y*, income	0.850	-0.055
Factor3	-0.114	1.054	*F*-index	0.827	0.040
Factor4	-0.120	1.000	*P*-index	0.526	0.103

Run for N = 1,965 overlapping observations. Gray shading indicates results that are of no consequence.

2.3 The problematic regression techniques

The toolkit of economists is full of regression techniques. For the study of transitions, they have four problems: (a) Transitions are slow, so the data have much inertia; (b) they occur in many variables, creating confluence; (c) they are non-linear; and (d) they are fuzzy. The present concentrates on problem (a).

The (X, y)-scatter often looks as one of the graphs on Figure 1.1, where the two variables are highly correlated. Typically, the data are short in time (like 2-3 decades) and wide in countries (like 150). Assume we try to estimate the effect of income with models (1) and (2):

(1) $X_{it} = \alpha + \beta_1 y_{it} + u_{1it}$ is a cross-country long-run relation, with residuals u.

(2) $X_{it} = \alpha + \gamma X_{it-1} + \beta_2 y_{it} + u_{1it}$ is the corresponding short-run adjustment model.

Regression (1) will typically get a positive coefficient, with a high *t*-ratio such as 20, so all looks well. If the short-run relation is clear and direct and the data are plentiful, (2) may also hold rather well, and it can be solved for the steady state.[1] It gives credibility if (1) and (2) give consistent estimates of the steady state slope β^*:

(3) $\beta^* \approx \beta_1 \approx \beta_2/(1 - \gamma)$

Provided the equivalence hypothesis is valid, β^* is an estimate of the slope of the transition curve, and the kernel curve discussed in the next section will have this slope. If X_{it} has too much inertia, γ may be so close to the unit root of one that (2) becomes shaky, and (3) breaks down.

The standard way to concentrate on the short run is to clean the relation for country differences and common time trends, which is done by breaking the constant into fixed effects α_i and α_t for countries and years respectively:

(4) $X_{it} = \alpha_i + \alpha_t + \beta y_{it} + u_{2it}$ by adding the short-run adjustments, it becomes

(5) $X_{it} = \gamma X_{it-1} + \alpha_i + \alpha_t + \beta y_{it} + u_{2it}$ L2FE-model (Lagged X, 2 fixed effects)[2]

If X_{it} has much inertia, α_i and X_{it-1} will be collinear, and hence the estimates of the fixed effects, notably α_i, may come to reduce the effect of X_{it-1}, making the estimate of γ too small, so that (3) does not hold any more. This is surely the case when institutions are stable most years and only change by occasional jumps, as the *P*-index (Polity). In such cases, (5) is the wrong tool; see the second part of Chapter 5.

Acemoglu *et al*. (2008) used model (5) to show that income has no effect on democracy, i.e. the *P*-index. Gundlach and Paldam (2010) replicate their result and show that in addition, it makes the Agricultural Transition and the Demographic Transition (from Figure 1.1) and several other transitions go away. This suggests that the L2FE-estimation model is the wrong tool for a strong and fuzzy relation with much inertia. As we are close to a unit root, anything

1. The steady state solution to (2) is reached by setting $X_{it-1} = X_{ti}$, and solving for X_{it}.
2. In the econometric literature, it is often termed a GDPM regression, for generic dynamic panel model.

can happen; see Chapter 5.5.

Transitions often have larger movements than necessary, i.e., the *G*-ratio is larger than 1. Thus, the variable fluctuates around the transition path. If it is below the path, an income shock will move it upwards, but if it is above the path, it may move it downwards. Hence, it is no wonder that the L2FE-estimate tends to find very little – even when there is a lot to be found.

2.4 Estimating the transition curve, Π: From scatters to kernels

The best method I have found to estimate transitions is to use kernel regressions on unified data sets organized by income. Such datasets easily become large, and the kernel shows the transition curve rather neatly.

The scatter-plot (X, y) of the data is a wide swarm of points that cover countries with much heterogeneity, but the raw scatter normally suggests a non-linear underlying curve as seen on Figure 1.1, and thus it averages into a neat curve by a kernel regression. If the curve has (most of) the properties listed in Table 2, it is interpreted as a transition curve:

(6) $X = \Pi(y) + u,$ where u is the noise term. (6) is estimated by the kernel

(7) $X = K^X(y, bw) \approx \Pi(y),$ where bw is the bandwidth

A kernel regression is a smoothed MA-process with a fixed bandwidth, *bw*. This book always estimates kernels by the Stata command *lpoly*. The program (and presumably all other kernel-programs) has a number of options. Two defaults are always used: The smoothing formula is Epanechnikov's kernel. The results are amazingly robust to variation of this choice. The degree of polynomial smooth is kept at zero. In addition, the following options are used: *nosc*atter suppresses the scatter, *ci* provides 95% *c*onfidence *i*ntervals, and *gen*erate gives the kernel curves as a data series.

Economic theory often predicts the qualitative form of a relation, as e.g. the predictions in Table 2. The confidence intervals allow us to test if a curve of the predicted form is possible within the intervals estimated. Large unified datasets normally yield amazingly narrow *ci*s, so the 'form test' is strong.

The robustness of the kernel is always analyzed by a set of experiments: Kernels are calculated for separate decades, and for different groups of countries. Systematic experiments are always made with the bandwidth, *bw*. Too short bandwidths give a wobbly curve, where some unexplainable stochastic fluctuations remain. As the bandwidth increases, the curve becomes smoother, but also gradually more linear, and in the end, it converges to a horizontal line at the average. In my experience, there always is a broad *bw*-range where the curve has the same form. The ‘central’ estimate is the curve with the clearest form. Deviations of 25% or even 50% in the *bw*, from the central estimate, are barely visible on the curve. Stata calculates a rule-of-thumb bandwidth, which is a fine starting point, but the weighting formula favors narrow *ci*’s that catch too many random fluctuations of the curve. I prefer interpretability, so a slightly larger bandwidth is preferred.

The great advantage of kernel regression is that it is a non-parametric technique, which does not presume a functional form. Thus, if the variable has a transition, the kernel will find a nice and often even a beautiful curve looking as Figure 1, with most or even all of the characteristics listed in Table 2. More advanced statistical methods exist – also as regards non-parametric techniques.[3] However, I am looking for the big pattern in the data, and I do not try to explain everything.

2.5 Scrambling tests in the data of the Main sample for the average kernel

Kernel regression $X = K^X(y, bw)$ starts by sorting all (X, y)-observations by income, y. The kernel is a useful analytical tool if the sorting randomizes the panel structure. This means that the countries and years do not cluster within the bandwidth. To examine if this is the case, two *scrambling tests* are run. The scrambling is also important to break the complex processes in the residuals, as further discussed in Chapter 8. All kernels presented use (subsets of) the 7,142 income observations in the Main sample. It spans 140 countries and 57 years from 1960 to 2016. The average kernel uses app 250 to 400 sorted observations from this data set. When it is divided by 325, it yields 22 non-overlapping subsets, which corresponds to the number of observations within the bandwidth of the average kernel.

Scrambling test 1: In the average subset, an observation, y_{it}, is followed by an observation for the same country, y_i, in 4.3% percent of the cases, and by an observation for the following year, y_{t+1}, in 3.5% of the cases. In 0.7% of the cases, y_{it} is followed by an observation for the same country and the following year, y_{it+1}. The scrambling is least complete in the first and the last of the 22 subsets. The three numbers listed fall to 2.9%, 3.2% and 0.5%, respectively, if the two utmost subsets are disregarded. I consider these results satisfactory.

Scrambling test 2: It calculates the country and time range in the average sample. The average number of countries within each subset is 52, or 37% of the 140 countries. Nearly all

[3] One of the papers received a referee report that concentrated on this point and recommended a handful of papers using frontline non-parametric statistics. These papers were indeed advanced, but the statistical techniques had managed to squeeze out all economics, and they reached results that neither the authors nor we could interpret. It is important that new methods are developed, but the development is only done when they become useful.

samples contain an observation that spans the full 57 years. It is only the last two subsamples that have a shorter time-range. Once again, this is rather satisfactory.

2.6 The attraction mechanics, the tension variable, Θ, and the excess movement

Section 1.3 in Chapter 1 discussed how transition curves should be understood. It suggested a bit of mechanics: If the kernel $X = K^X(y, bw)$ shows a nice $\Pi^X(y)$-curve, it is the hypothetical equilibrium path during the transition. There are many reasons for the big variation around the Π^X-curve, but there must be a pull from the curve. It means that the Π^X-curve is an ***attractor*** for X. All kinds of disturbances occur to push X into (random?) walks around the Π^X-path, but X is also pulled toward the Π^X-curve. I propose that the pull from the attractor $\Pi^X(y)$ is proportional to the distance Θ (theta) from the transition curve:

(8a) $\Theta = X - \Pi^X(y)$ and (8b) $\Delta X = -\alpha\,\Theta$

Θ is termed the ***tension*** variable (for X). If Θ is positive, the country has too much X at its level of income, and I predict that the country will come to see a fall in X. If Θ is negative, the country has too little X at its level of income, and I predict that the country will come to see an increase in X. Thus, Θ and ΔX should be negatively correlated.

If X is on the equilibrium path, i.e. $X = \Pi^X(y)$, $\Theta = 0$. However, when income increases from y_0 to y_1, while X is constant, X moves a little away from equilibrium, as can be seen as a change in the tension. The change is: $\Delta\Pi^1 = \Pi(y_1) - \Pi(y_0)$, so the tension changes from Θ^0 to $\Theta^1 = \Theta^0 + \Delta\Pi^1$.

Think of the Democratic Transition. If a country has too much democracy (Θ^0 is positive), economic growth will reduce the tension. Vice versa, if a country has too little democracy (Θ^0 is negative), economic growth will increase the tension. If X is not on the Π-path, we expect an adjustment of X toward $\Pi(y)$ as a movement in n steps of Θ from the original value Θ^0. Step 1 reduces Θ from Θ^0 to Θ^1.[4]

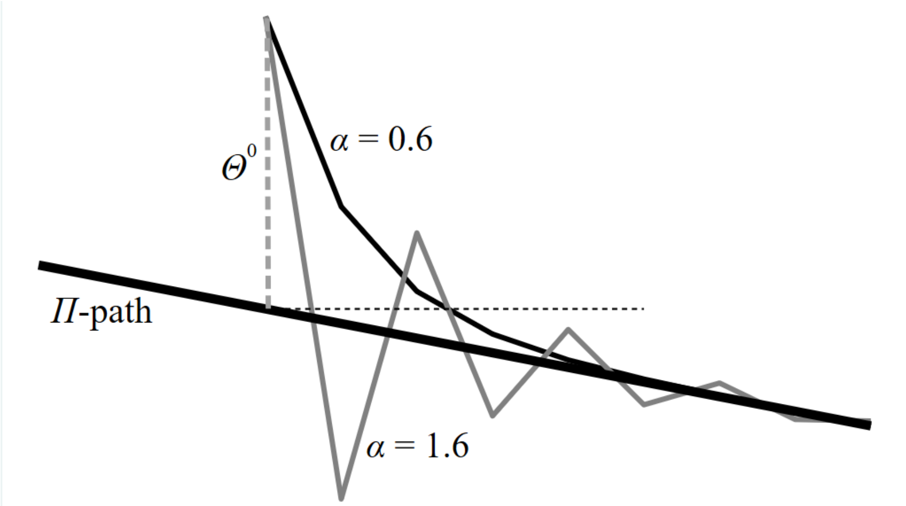
(9) Step 1: $\Theta^1 = (1-\alpha)\,\Theta^0$ during step 1 income grows, so Π changes by $\Delta\Pi^1$
Step 2: $\Theta^2 = (1-\alpha)^2\,(\Theta^0 + \Delta\Pi^1)$ this continues
Step n: $\Theta^n = (1-\alpha)^n\,(\Theta^0 + \Delta\Pi^{n-1})$ where $\Delta\Pi^{n-1}$ is the change in Π since the start

Equation (9) makes sense only if α is in the interval]0, 2[, so that numerically the tension falls

[4] Chapters 5 and 6 show that while the adjustments are as predicted, it is very difficult to predict when they occur.

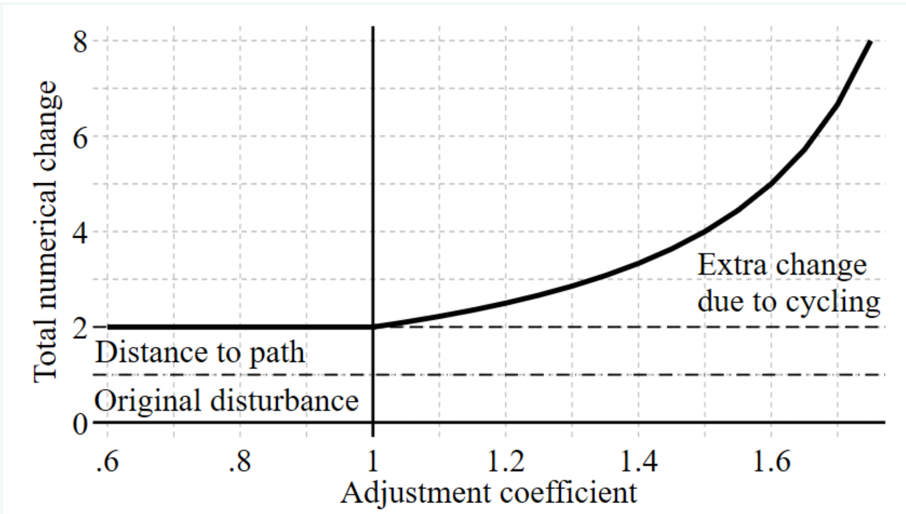
if $\Delta\Pi$ is small (as it surely is). If the steps are small, such as a year, $\Delta\Pi$ is negligible, but (as showed in Chapter 7) the steps might be long such as a decade apart. Thus, $\Delta\Pi$ might not be negligible. Consider a situation where X is on the Π-path. Now one disturbance occurs that pushes X to Θ^0 – no further disturbances occur. As the number of adjustments n rises, the numerical value $|\Theta^n| \to 0$. Thus, X converges to the Π-curve, as shown on Figure 2.

Figure 2. The effect of the disturbance Θ^0 and the resulting adjustments



Π-path for α = 0.6 is the black declining transition path on the Figure. The eight first steps of the adjustment process are visible. For α = 1.6, it gives the gray adjustment path, where every step overshoots the *Π*-path. The horizontal dotted line shows when the adjustment has reached the old value of *X*, but in the meantime, *Π* has moved due to the growth of income.

Figure 3. The total numerical change as function of α, the adjustment coefficient



To the left of the vertical line (at α = 1), *X* returns to the equilibrium path by steps that add to 1. Thus, the total change is $2\Theta^0$. The size of α determines the adjustment speed. To the right of the vertical line, the adjustments are damped cycles, giving extra numerical changes. More cycles are needed if α is large.

If α is in the interval]0, 1], the adjustments sums to Θ^0, and the number of steps necessary to go back to the Π-path is larger, the smaller α is. If $\alpha = 1$, X goes to the path in one step. The total amount of numerical change in X due to the disturbance is thus $2\Theta^0$.

If α is in the interval]1, 2[, the signs on Θ^n change from step to step in the process, as shown with the gray zigzag curve on Figure 2. This means the *X overshoots* the Π-path, so that the convergence to the Π-path takes the form of damped oscillations around the path. Hence, the sum of the numerical steps is larger than the original disturbance.

Figure 3 suggests the size of the total variation caused by a random disturbance of 1 unit. If $\alpha = 1.5$ (as found in Chapter 5), the total change becomes 4. A disturbance of one *P*-point leads to three extra *P*-points of changes in due time. This is all parts of the excess amount of institutional change leading to the large *G*-ratios found in Chapter 14. While the mechanism of the adjustment to the path is simple, the timing of the steps is a more difficult question, which will be a recurrent theme in Part II of the book.

2.7 Three ways to establish the main causal direction

I now turn to methods for distinguishing between two causal directions: $y \Rightarrow X$ and $X \Rightarrow y$. This section discusses three methods, while section 8 discusses a fourth one.

Method one. An economic model is a causal explanation. If it fits the data, this is causal evidence; see in particular Chapter 5 on the Jumps Model. In addition, I use three statistical tools to get an empirical handle on causality.

Method two. Causality in the long run. The ***beauty test*** for kernels (see Paldam 2019). Transition-theory proposes that $y \Rightarrow X$, and that $X = \Pi^X(y)$ has the form described in Table 2. Thus, the kernel $X = K^X(y)$ should have the said form. If an alternative Λ-theory proposes that $X \Rightarrow y$, and that $y = \Lambda(X)$ has a particular form, it can be estimated by $y = K^y(X)$. Hence, we compare the two kernel estimates that correspond to two theories:

(10) $X = K^X(y, bw) \approx \Pi(y)$ which tells us if the Π-theory can explain the data

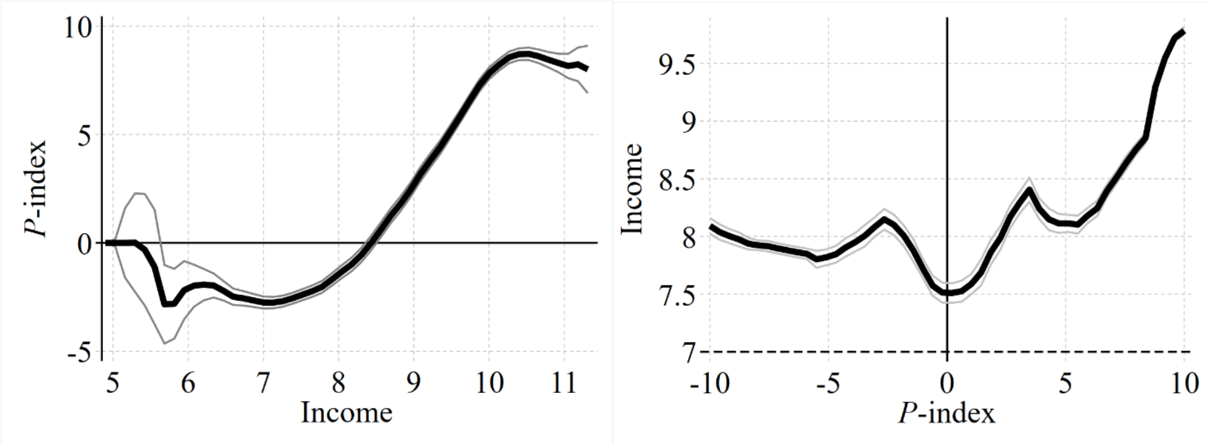
(11) $y = K^y(X, bw) \approx \Lambda(X)$ which tells us if the Λ-theory can explain the data

In (10) the dataset is sorted by y, and the average (over bw) is calculated around each y. In (11) the dataset is sorted by X, and the average (over bw) is calculated around each X. These calculations are quite different, so it is no wonder that the reverse kernels often look strikingly different. Furthermore, normally only one looks as it should by its theory. This is causal evidence in favor of the winning theory. The two graphs on Figure 4 show the Democratic

Transition from Chapter 4 and the reverse curve. The Transition curve looks beautiful, while the reverse makes little sense. Thus, Figure 4 (strongly) suggests that the main direction is from income to the *P*-index. The Democratic Transition (shown) and the Transition of Corruption analyzed in Chapters 4 and 10, respectively, both lead to beautiful curves, while the reverse kernels do not look as they should by any theory.

Figure 4. Example of two reverse causal theories

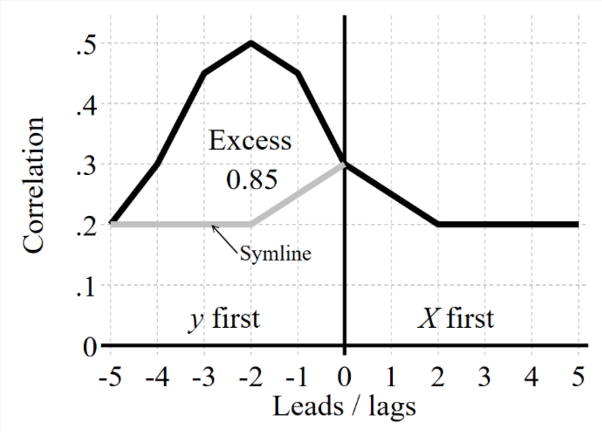
Figure 4a. Income explains political system Figure 4b. Political system explains income



Both curves use the default bandwidth chosen by Stata. The income data are thin below 6.4. The economics of the curves is discussed in Chapter 4.

Method three. Causality in the short run. Most chapters use a Granger-like causality test in the form of correlograms based on time-series of n_1 years for n_2 countries. For country *i* the correlogram is:

Figure 5. Hypothetical example of correlogram test



(12) cor(X_{it}, y_{it+j}), where $j = -q, \dots, +q$, where q is a number such as 5 or larger

(13) Each point drawn as the bold black curve is the average of the n_2 country-correlations

Figure 5 is a hypothetical illustration of the average correlogram. Long-run confluence appears as a level of the correlogram. On the figure, the underlying level is 0.2. Correlograms often have a hump that goes above or below that level. If the hump is for *X* before *y*, it means that *X* is a predictor of *y*. If the hump is for *y* before *X*, then *y* is a predictor of *X*, as in the example. It indicates that *y* predicts *X*, and is taken as an indication of the causal direction. To see the hump as clearly as possible, the gray *symline* is drawn. It gives the half of the correlogram that has the lowest correlations drawn symmetrically around the vertical axis.

If the two variables correlated are simultaneous, the correlogram is symmetric around a peak at zero, drawn as the vertical axis. Asymmetry suggests causality. It is indicated as the *excess* area on the figure. It is also given as a sum of the excess for each lead. The example is 'unusually' neat. The symline and the correlogram will normally start at different points, the two curves may intersect more than once, and the underlying level may have a clear trend etc.

Each average correlation indicated as a kink-point for the correlogram is calculated for n_1 years. This allows a test of significance for (H0: cor = 0). The two-sided 5% level of significance for the correlation for $n_1 = 50$ is cor* = 0.27. Most of the correlograms are averages of such correlograms for n_2 countries. If the country observations were independent, the aggregate level of significance should be cor*/$\sqrt{n_2}$. If n_2 is 100, this reduces cor* by a factor 1/10. However, the country observations are dependent, so the reduction is not so large. Below I use a crude rule of thumb that the correlations in the correlograms are significant if they are numerically larger than 0.09. In some cases, the correlograms are rather different across countries and then it is better to calculate the standard error of the correlation with the same lag (see Figure 6.11 in Chapter 6).

2.8 The fourth way: The DP-tests –using the development potential before it happened

Method fourth to establish long-run causality is the ***DP-test*** that uses variables for the ***d***evelopment ***p***otential of countries as instruments.[5] These variables measure the nature-given development potential long before modern development started – they are reported at the end of the section. While they predict the development of countries fairly well, it appears unlikely that they can predict the institutions of countries. As the *DP*-variables are time-invariant, the

[5] The DP-test is from Gundlach and Paldam (2009). It is used in Chapters 4, 8 9, 10 and 11.

test works only on cross-country samples. Table 4 gives the mechanics of the TVIV-test.

(15a) to (15c) use the instrument *D* to see if *y* causes *X*. However, first a set of tests has to be made to say if the instrument *D* is strong and valid. If the tests are accepted, and the estimate of β_2 is significantly different from zero, a causal link from *y* to *X* has been established. Furthermore, if β_1 from equation (14) and β_2 from equation (15c) are the same, we draw the strong conclusion that the causal relation from *y* to *X* is the only relation between *y* and *X*.

Table 4. The equation of the standard TSIV-test

Transition causality		Reverse causality		Comment
(14)	$X_i = \alpha_1 + \beta_1 y_i + u_{1i}$	(16)	$y_i = \alpha_3 + \beta_3 X_i + u_{4i}$,	Simple regression of *X* and *y*
(15a	$y_i = \gamma_2 + \lambda_2 D_i + u_{2i}$	(17a)	$X_i = \gamma_2 + \lambda_2 D_i + u_{5i}$	*D* is the instrument
(15b)	$y_i^D = \gamma_2 + \lambda_2 D_i$	(17b)	$X_i^D = \gamma_2 + \lambda_2 D_i$	Calculate the instrumented variable
(15c)	$X_i = \alpha_2 + \beta_2 y_i^D + u_{3i}$	(17c)	$y_i = \alpha_4 + \beta_4 X_i^D + u_{6i}$	The TSIV estimate

The six *u*-variables are the residual terms.

The strong conclusion should mean that reverse relations (17a) and (17c) detect no causality. This is often discovered already when the tests show that *D* is not a strong and valid instrument for *X*. However, a couple of cases are found where (weaker) reverse causality is detected. In this case the (*y*, *X*)-relation has simultaneity.

As the test deals with the long run, the DP-tests run use averages over the period 2005-10 if the data are annual.[6] These data are unlikely to be revised, so the test should replicate nicely. The regressions (14) and (15) are run for one year, but cross-country samples exist for about 40 years, so the regressions can be run every year, and the 40 sets of coefficient estimates can be used to test the robustness of the results. The DP-test is run for all seven institutional variables listed in Table 1.3. All TSIV regressions report four tests:

The ***Cragg-Donald*** (CD) test for instrument strength. If it is below the critical value (10 percent maximal size), the instruments are weak. The critical value is between 20 and 22. Thus, if the CD-value reported exceeds the critical value, I say that the instruments are strong, and the test values are bolded.

The ***Sargan*** test for overidentification reject the joint null hypothesis that the instruments are valid and correctly excluded from the estimate. Here the p-value is reported; it should show that the test is not rejected, i.e. the p-values are above 0.05, preferably above 0.15.

[6] The period includes both 4 very good years and 2 crises years, so they are about average.

Table 5. The DP-variables

	The biological variables:
Animals	Number of domesticable big mammals, weighing more than 45 kilos, which are believed to have been present in various regions of the world in prehistory.
Plants	Number of arable wild grasses known to have existed in various regions of the world in prehistory, with a mean kernel weight exceeding 10 mg.
Bioavg	Average of plants and animals, where each variable was first normalized by dividing by its maximum value.
Biofpc.	The first principal component of plants and animals.
Maleco	Measure of malaria ecology. It combines climatic factors and biological properties of the regionally dominant malaria vector into an index of the stability of malaria transmission (malaria ecology). The index is an average for each country of highly disaggregated sub-national data. Source: Kiszewski *et al.* (2004).
	The geographic variables:
Axis	Relative east-west orientation of a country, measured as east–west distance (longitudinal degrees) divided by north–south distance (latitudinal degrees).
Climate	A ranking of climates according to how favorable they are to agriculture, based on the Köppen classification.
Coast	Proportion of land area within 100 km of the seacoast. Source: McArthur and Sachs (2001).
Frost	Proportion of a country's land receiving five or more frost days in that country's winter, defined as December through February in the Northern hemisphere, and June through August in the Southern hemisphere. Source: Masters and McMillan (2001).
Geoav	Average of climate, lat, and axis, where each variable was first normalized by dividing by its maximum value.
Geofpc	The first principal component of climate, *lat*, *axis* and *size*.
Lat	Distance from the equator as measured by the absolute value of country-specific latitude in degrees divided by 90 to place it on a [0,1] scale. Source: Hall and Jones (1999).
Size	The size of the landmass to which the country belongs, in millions of square kilometers (a country may belong to Eurasia or it may be an island).

Variables reported without source are from Hibbs and Olsson (2004) and Olsson and Hibbs (2005). To include Ethiopia in the 1995 sample, the 1993 observation for *polity* is used. Belize, Cap Verde, Hong Kong, Iceland, Luxembourg, Maldives, Malta, and Samoa are not included in the Polity IV database. Fiji, Papua New Guinea, and the Solomon Islands are not included in the Maddison database. The estimation results are not statistically significantly affected by the additional observation on Ethiopia.

The ***Hausman*** test for parameter consistency of OLS and IV estimates, i.e., does β_1 differ significantly from β_2. Once again, the test tries to reject homogeneity, so the p-values should be above 0.05, preferably above 0.15.

The last section in the tables reporting the DP-tests is a test for the detection of reverse causality. It runs Equations (16) and (17), explaining *y* by *X*. Here I just report the Cragg-Donald test, which in all cases shows that the instruments are weaker in this case, but sometimes they are acceptable, indicating some simultaneity.

The DP-variables are given in Table 5. The idea of the DP-variables and most of the effort to put the variable together are due to D.A. Hibbs and O. Olsson. They use the suggestions of Jared Diamond (from Diamond 1997). The variables are to represent facts that predict development before it happened. As development has long roots, they should be from before these roots. The variables are biological and geographical. Hibbs and Olsson have compiled the variables reported without source. The biogeography data include 112 country observations. If income data or institutional index data are missing for 1995 (or another of the selected cross section years), the next observation within a time interval of +/−10 years is used.

In regressions explaining present day income, these variables work rather well. The four overseas western countries (Australia, Canada, New Zealand and the USA) were established in areas with rather low values as regards *animals* and *plants*. The largest domesticable animal in North America was the turkey, and there were no such animals in Australia and New Zealand. However, the European immigrants quickly brought such animals and plants. Thus, the four overseas Western countries may be treated as part of Europe.

2.9 Can transition curves be statistical artifacts?

The following section does not come naturally anywhere in the book. It discusses an argument I have sometimes encountered. Maybe *Π*-curves are an artefact due to the definition of the institutional indices as integers in a closed interval. The argument exists in two versions exemplified by the Polity index *P* that is defined on [−10, +10].

The first version is the *truncation argument*: It claims that a true index measuring the political system would use a larger scale; thus, the bends at the top and the bottom have been created by limiting the possibilities to the closed set of integers. The problem is small at the bottom, as the flat part of the curves is some distance from the end of the interval, but the *Π*-curve does converge to the top of the interval. This may be an artefact.

My assessment is that it is no artefact, but a real fact. There is little evidence that mass democracy can be more democratic than it is in the typical developed Western country – proposals for more democracy are either marginal or utopian. The author has participated in a study comparing Denmark and Switzerland. The two countries both had top scores on all three indices (*P*, *CL* and *PR*), and in the World Values Survey, a large part of the respondents in both countries answered that the country had full democracy. However, the two countries have rather different political institutions and traditions; see Chapter 6 in Christoffersen *et al*. 2016. It is possible that the best institutions of the two countries could be combined into a more democratic system, but it is more likely that a combined system would be too complex to operate. Thus, I think that the upper bound is real.

The other argument is that the transition path can be modeled as a *random walk in a closed interval*. A pure random walk does not generate a *Π*-curve. It will inevitably be trendless around the center of the interval (which is zero for *P*). However, if a drift toward the *Π*-curve is added, then it can be made to work, but then we have reached a version of the Jumps Model that is discussed in Chapter 5.