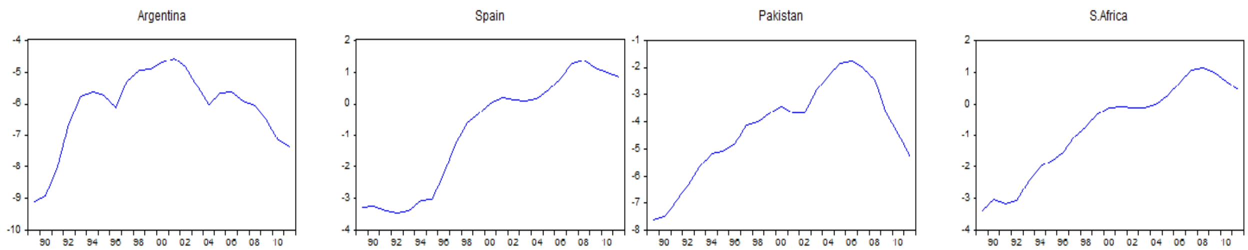
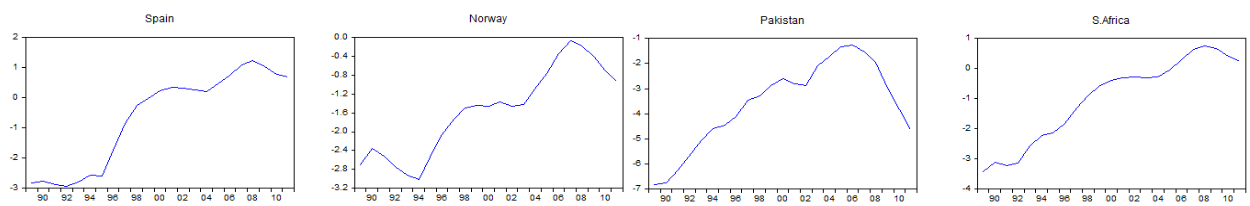
Financial Development, Structure and Growth: New Data, Method and Results
Online Appendix

Figure 1: Plots of selected time series for selected countries.

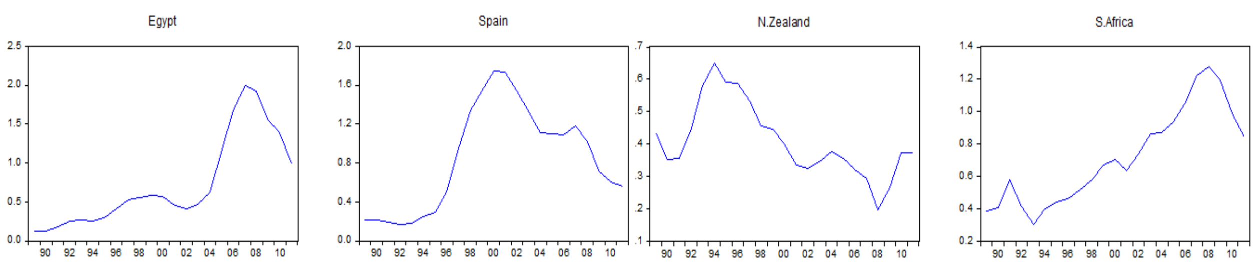
F^{AG} for Selected Countries



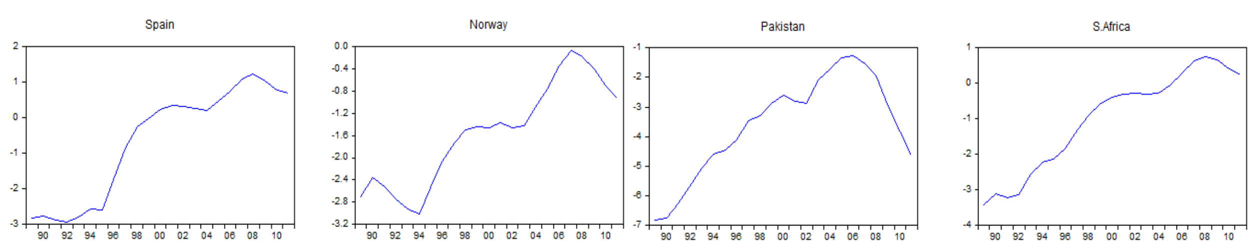
S^{AG} for Selected Countries



F^{A} for Selected Countries

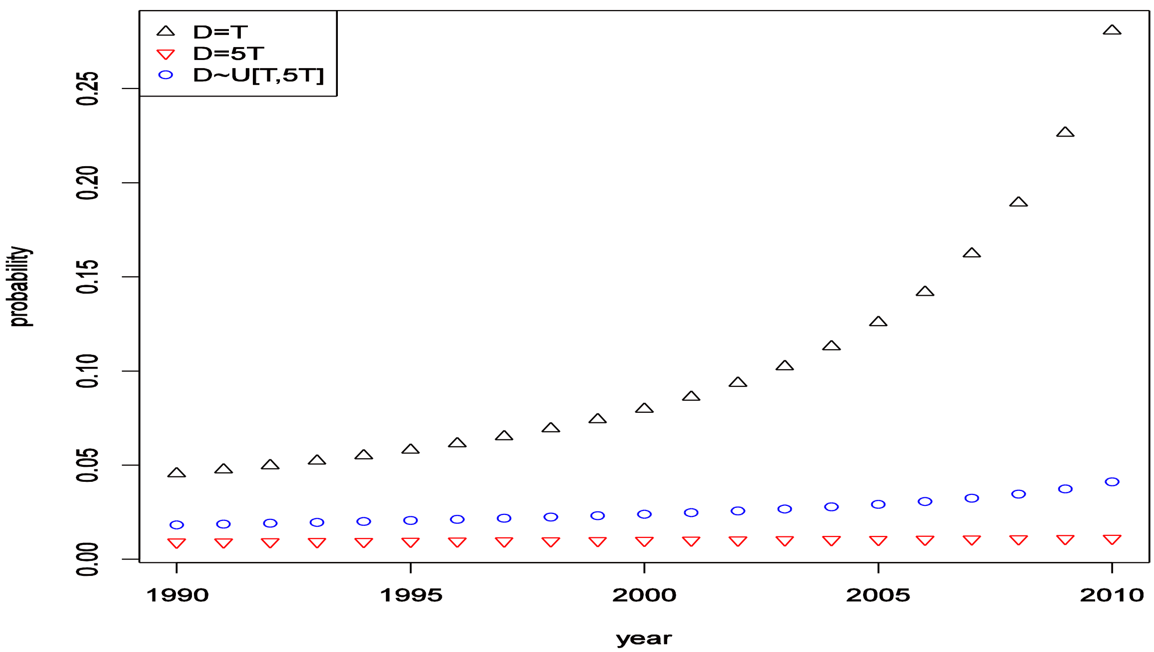


S^{A} for Selected Countries



F^{AG}, S^{AG}, F^{A} and S^{A} respectively stand for finance aggregate, structure aggregate, finance activity and structure activity as defined in the notes to Table 3 in the paper.

Figure 2: The priors for a particular period being a break unconditional on the previous periods for different specifications of D when T=22



The vertical axes shows the probability $Pr(B_t = 1)$ unconditional on previous periods. It is represented with the symbol △ when D=T and in this case the prior probability of a break significantly increases with time. When D=5T the prior probability of a break (represented with ▽) remains roughly constant over the whole sample period. In our empirical analysis we specify a uniform distribution in (T, 5T) as a prior for D, such that there is only a slight increase in the prior probability of a break over time (as shown in the graph with the symbol o).

Table A1: Identification of the order of lags, leads and the deterministic components

		With breaks					no breaks				
Case		Δx_t	$L\Delta x_t$	$L^{+}\Delta x_t$	trend	LPL	Δx_t	$L\Delta x_t$	$L^{+}\Delta x_t$	trend	LPL
High-Income-Panel	M1	0	0	0	quadratic	475.879	0	0	0	quadratic	475.849
	M2	0	0	0	quadratic	474.205	0	0	0	quadratic	474.094
	M3	0	0	0	quadratic	481.865	0	0	0	quadratic	481.843
	M4	0	0	0	quadratic	479.997	0	0	0	quadratic	479.959
	M5	0	0	0	quadratic	474.174	1	0	0	quadratic	474.545
	M6	0	0	0	quadratic	480.432	0	0	0	quadratic	480.416
	M7	0	0	0	quadratic	472.225	0	0	0	quadratic	471.795
	M8	0	0	0	quadratic	483.329	0	0	0	quadratic	483.31
	M9	1	0	0	quadratic	487.872	1	0	0	quadratic	482.111
Middle-and-Low-Income-Panel	M1	1	0	1	quadratic	515.413	1	0	1	quadratic	515.413
	M2	1	0	1	quadratic	516.066	1	0	1	quadratic	516.068
	M3	1	0	1	quadratic	510.099	1	0	1	quadratic	510.098
	M4	1	0	1	quadratic	525.447	1	0	1	quadratic	525.529
	M5	1	0	1	quadratic	524.622	1	0	1	quadratic	524.621
	M6	1	0	1	quadratic	520.706	1	0	1	quadratic	520.877
	M7	1	0	1	quadratic	528.244	1	0	1	quadratic	528.244
	M8	1	0	1	quadratic	524.735	1	0	1	quadratic	524.788
	M9	1	0	1	quadratic	517.057	1	0	1	quadratic	517.056
Full-Panel	M1	1	0	0	quadratic	992.417	1	0	0	quadratic	992.416
	M2	1	0	0	quadratic	991.15	1	0	0	quadratic	991.146
	M3	1	0	0	quadratic	990.524	1	0	0	quadratic	990.523
	M4	1	0	0	quadratic	1001.024	1	0	0	quadratic	1001.024
	M5	1	0	0	quadratic	990.744	1	0	0	quadratic	990.743
	M6	1	0	0	quadratic	993.985	1	0	0	quadratic	994.029
	M7	1	0	0	quadratic	985.209	1	0	0	quadratic	985.212
	M8	1	0	0	quadratic	1002.025	1	0	0	quadratic	1002.019
	M9	1	0	0	quadratic	990.964	1	0	0	quadratic	990.964

Note: This Table reports the log of predicted likelihood (LPL) for our empirical specifications (DOLS). Each reported specification is based on the highest value of the LPL. We search through the first order leads and lags and the deterministic components consisting of a constant, a linear and a quadratic trend. This generates 15 potential specifications to exhaust all possible combinations of leads ($L^{+}\Delta x_t$), lags ($L\Delta x_t$) and contemporaneous first differences terms (Δx_t) of I(1) regressors and the deterministic components. In all specifications, all three deterministic components – a constant, a linear and a quadratic trend – appear significant, showing the highest predictive likelihood hence maintained.

Table A1 reports the highest value of the log of predictive likelihood computed across all 15 specifications to exhaust all possible combinations of leads, lags and the deterministic components for each specification. The results show that the empirical specifications of DOLS regressions differ across panels and models. When models do not allow for breaks, the DOLS specifications for the full panel require augmentation only by the contemporaneous first differences of I(1) regressors; no lead and lag augmentations are required. The panel of high-income countries does not require any augmentations except in two cases (Models 5 and 9) – these two requiring augmentation by contemporaneous difference terms only. In contrast, the panels of middle-and-low-income countries require augmentations by contemporaneous and lead terms of first differenced covariates. Allowing for potential breaks does not alter the empirical specifications except for only one case (Model 5 for high-income countries).

Table A2: Estimation results with structural breaks for high-income countries in selected years.

Models	Year	k	F^{AG}	S^{AG}	S^{A}	S^{Z}	F^{Z}	F^{A}
M1	1990	0.924 (0.796,1.05)	0.0058 (-0.006,0.013)	0.0175 (-0.054,0.026)	–	–	–	–
	2008	0.923 (0.796,1.049)	0.0058 (-0.002,0.013)	0.0186 (0.011,0.027)	–	–	–	–
	2009	0.922 (0.792,1.049)	0.005 (-0.006,0.013)	0.0189 (0.001,0.030)	–	–	–	–
M2	1990	0.918 (0.79,1.043)	0.0061 (-0.005,0.013)	–	0.0151 (-0.121,0.026)	–	–	–
	2008	0.917 (0.789,1.043)	0.0061 (-0.0008,0.01)	–	0.0163 (0.006,0.026)	–	–	–
	2009	0.914 (0.786,1.042)	0.0062 (-0.003,0.015)	–	0.0091 (-0.02,0.027)	–	–	–
M3	1990	0.903 (0.78,1.028)	0.0065 (-0.0007,0.013)	–	–	0.0343 (0.020,0.045)	–	–
	2008	0.903 (0.78,1.028)	0.0064 (-0.0005,0.013)	–	–	0.0345 (0.023,0.045)	–	–
	2009	0.901 (0.775,1.027)	0.0057 (-0.007,0.013)	–	–	0.0376 (0.021,0.068)	–	–
M4	1990	0.924 (0.802,1.046)	–	0.0170 (-0.013,0.024)	–	–	0.0220 (0.008,0.033)	–
	2008	0.924 (0.801,1.046)	–	0.0172 (0.01,0.024)	–	–	0.0220 (0.009,0.033)	–
	2009	0.924 (0.801,1.046)	–	0.0171 (0.002,0.025)	–	–	0.0221 (0.007,0.034)	–
M5	1990	0.923 (0.798,1.049)	–	0.0157 (-0.065,0.026)	–	–	–	0.0029 (-0.010,0.011)
	2008	0.921 (0.797,1.046)	–	0.0193 (0.011,0.027)	–	–	–	0.0032 (-0.004,0.011)
	2009	0.919 (0.794,1.045)	–	0.0194 (0.0002,0.031)	–	–	–	0.0017 (-0.008,0.011)
M6	1990	0.925 (0.802,1.044)	–	–	0.0166 (0.006,0.026)	–	0.0249 (0.011,0.036)	–
	2008	0.925 (0.801,1.044)	–	–	0.0167 (0.007,0.026)	–	0.0249 (0.011,0.036)	–
	2009	0.924 (0.8,1.044)	–	–	0.0159 (-0.018,0.026)	–	0.0261 (0.010,0.038)	–
M7	1990	0.914 (0.787,1.04)	–	–	0.0153 (-0.175,0.026)	–	–	0.0047 (-0.005,0.013)
	2008	0.913 (0.787,1.039)	–	–	0.0169 (0.007,0.027)	–	–	0.0044 (-0.002,0.011)
	2009	0.91 (0.783,1.036)	–	–	0.0043 (-0.023,0.027)	–	–	0.0061 (-0.004,0.017)
M8	1990	0.909 (0.787,1.031)	–	–	–	0.0318 (0.019,0.043)	0.0185 (0.003,0.030)	–
	2008	0.909 (0.787,1.031)	–	–	–	0.0319 (0.021,0.043)	0.0184 (0.003,0.030)	–
	2009	0.908 (0.786,1.031)	–	–	–	0.0327 (0.017,0.056)	0.0187 (-0.003,0.030)	–
M9	1996	1.01 (0.895,1.124)	–	–	–	0.1148 (0.093,0.136)	–	-0.0194 (-0.028,-0.011)
	2005	1.016 (0.901,1.131)	–	–	–	0.0292 (0.018,0.040)	–	-0.0087 (-0.015,-0.002)
	2006	1.016 (0.901,1.131)	–	–	–	0.0296 (0.019,0.045)	–	-0.0088 (-0.016,-0.003)

Note: The numbers inside () indicate the 95% posterior credible intervals. Variables and specifications are given in the notes to Table 4 of the paper.

2. Simulation Details

Denote $x_{it} = (k_{it}, fd_{it}, fs_{it})'$ and $X_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$, which is a $T \times 3$ matrix. For each cross-sectional unit i, stack up all the observations of T periods to obtain,

$$\boldsymbol{y_i} = \boldsymbol{W_i\gamma_i} + \boldsymbol{X_i\beta_1} + \boldsymbol{\Xi_i u} + \boldsymbol{\varepsilon_i} \tag{A1}$$

where,

$y_i = (y_{i1}, \ldots, y_{iT})'$, $\gamma_i = \left(\gamma_{1i}, \gamma_{2i}, \gamma_{3i}, \gamma_{4i,-d_1}, \ldots, \gamma_{4i,d_2}, \gamma_{5i,-d_1}, \ldots, \gamma_{5i,d_2}, \gamma_{6i,-d_1}, \ldots, \gamma_{6i,d_2}\right)'$; $u = (u_1', u_2', \ldots, u_{T-1}')'$; W_i is a matrix of T rows with the t[th] row being,
$\left(1, t, t^2, \Delta k_{i,t-d_1}, \ldots, \Delta k_{i,t+d_2}, \Delta fd_{i,t-d_1}, \ldots, \Delta fd_{i,t+d_2}, \Delta fs_{i,t-d_1}, \ldots, \Delta fs_{i,t+d_2}\right)$ for $t = 1, \ldots, T$; $\Xi_i = (B_1X_{1,i}, B_2X_{2,i}, \ldots, B_{T-1}X_{T-1,i})$ and $X_{j,i}$ is a matrix with the first j rows being 0 and the remaining rows being the same as those in X_i. The posterior of $p(\theta|y)$ is proportional to the product of the prior $p(\theta)$ and the likelihood function $p(y|\theta)$:

$$\boldsymbol{p(\theta|y) \propto p(\beta_1|\sigma^2)p(u|\sigma_u^2)p(\sigma_u^2)p\left(B_{1:(T-1)}\middle|D\right)p(D)} \tag{A2}$$
$$\boldsymbol{(\sigma^2)^{-\frac{NT-2}{2}} exp\left[-\frac{\sum_{i=1}^{N}(y_i - W_i\gamma_i - X_i\beta_1 - \Xi_i u)'(y_i - W_i\gamma_i - X_i\beta_1 - \Xi_i u)}{2\sigma^2}\right]}.$$

After integrating out γ_i from the posterior kernel, one can obtain:

$$\boldsymbol{p\left(\sigma^2, \beta_1, u, \sigma_u^2, B_{1:(T-1)}, D\right) \propto p(\beta_1|\sigma^2)p(u|\sigma_u^2)p(\sigma_u^2)p\left(B_{1:(T-1)}\middle|D\right)} \tag{A3}$$
$$\boldsymbol{(\sigma^2)^{-\frac{N(T-k_\gamma)-2}{2}} exp\left[-\frac{\sum_{i=1}^{N}(y_i - X_i\beta_1 - \Xi_i u)'M_i(y_i - X_i\beta_1 - \Xi_i u)}{2\sigma^2}\right]},$$

where k_γ is the number of elements in γ_i and $M_i = I_T - W_i(W_i'W_i)^{-1}W_i'$. We can then derive the posterior distribution of $(\beta_1, u, B_{1:(T-1)})$ conditional on $(D, \sigma^2, \sigma_u^2)$ as the following:

$$\boldsymbol{\beta_1|u, y, B_{1:(T-1)}, \sigma^2, \sigma_u^2 \sim N(V_\beta \textstyle\sum_{i=1}^{N} X_i' M_i(y_i - \Xi_i u), \sigma^2 V_\beta)}, \tag{A4}$$

$$\boldsymbol{u|\, y, B_{1:(T-1)}, \sigma^2, \sigma_u^2 \sim N(V_u \textstyle\sum_{i=1}^{N} \Xi_i' M_i X_i V_\beta \sum_{i=1}^{N} X_i' M_i y_i\,, \sigma^2 V_u)}, \tag{A5}$$

$$p\left(B_{1:(T-1)}\middle|y, D, \sigma^2, \sigma_u^2\right)$$
$$\propto p(B_{1:(T-1)}|D)|\sigma^2 V_u|^{\frac{1}{2}} exp\left[-\frac{\sum_{i=1}^{N} y_i' y_i - \sum_{i=1}^{N} y_i' M_i X_i V_\beta \sum_{i=1}^{N} X_i' M_i y_i}{2\sigma^2}\right] \tag{A6}$$
$$exp\left[\frac{\sum_{i=1}^{N} y_i' M_i X_i V_\beta \sum_{i=1}^{N} X_i' M_i \Xi_i V_u \sum_{i=1}^{N} \Xi_i' M_i X_i V_\beta \sum_{i=1}^{N} X_i' M_i y_i}{2\sigma^2}\right],$$

where $V_\beta = \left(\sum_{i=1}^{N} X_i' M_i X_i + \frac{1}{\tau} I_3\right)^{-1}$ and $V_u = \left(\sum_{i=1}^{N} \Xi_i' M_i X_i V_\beta \sum_{i=1}^{N} X_i' M_i \Xi_i + \frac{\boldsymbol{\sigma^2}}{\boldsymbol{\sigma_u^2}} I\right)^{-1}$. To sample $B_{1:(T-1)}$ from its conditional posterior, we set up another Gibbs sampler, which draws each element in $B_{1:(T-1)}$ conditional on the other elements. The posterior of $(D, \sigma^2, \sigma_u^2)$ conditional on $(\beta_1, u, B_{1:(T-1)})$ is:

$$p(D|B_{1:(T-1)}, y) \propto p(D)p(B_{1:(T-1)}|D) \tag{A7}$$

$$\sigma^2|\beta_1, u, B_{1:(T-1)}, y \sim IG\left(\textstyle\sum_{i=1}^{N}(y_i - W_i\gamma_i - X_i\beta_1 - \Xi_i u)'(y_i - W_i\gamma_i - X_i\beta_1 - \Xi_i u), N(T - k_\gamma)\right), \tag{A8}$$

$$\sigma_{uj}^2|\, u, B_{1:(T-1)}, y \sim IG\left(\textstyle\sum_{t=1}^{T-1} B_t u_{jt}^2 + 0.1, 3 + \sum_{t=1}^{T-1} B_t\right), for\ j = 1,2,3. \tag{A9}$$

Regarding the calculation of the predictive likelihood, Geweke (1995, 1996) shows that the predictive likelihood $p_{T_0}^{T}$ can be decomposed as the following product,

$$p_{T_0}^{T} = p_{T_0}^{T_0+1} p_{T_0+1}^{T_0+2} \dots p_{T-1}^{T} \tag{A10}$$

where each of the components of the predictive likelihood is a one-step-ahead predictive likelihood, which can be approximated by using draws from the posterior as:

$$\hat{p}_{T_0+t}^{T_0+t+1} = \frac{\sum_{s=1}^{S} p\left(y_{T_0+t+1} \middle| \theta_{T_0+t+1}^{s}, y_{1:(T_0+t)}\right)}{S}, for\ t = 0,1, \dots, (T - T_0 - 1). \tag{A11}$$

where $(\theta_{T_0+t+1}^{1}, \dots, \theta_{T_0+t+1}^{S})$ are S draws from the posterior of θ_{T_0+t+1} given $y_{1:(T_0+t)}$. For the model (1) in page 7, we just need to predict B_{T_0+t} and u_{T_0+t} after the estimation with the initial sample up to period $T_0 + t$. The posterior of θ_{T_0+t+1} given $y_{1:(T_0+t)}$ for our model is

$$p\left(\theta_{T_0+t+1} | y_{1:(T_0+t)}\right) = p\left(u_{T_0+t} \middle| B_{T_0+t}, \sigma_u^2, y_{1:(T_0+t)}\right) \\ p(B_{T_0+t} | B_{1:(T_0+t-1)}, D, y_{1:(T_0+t)}) p(D, \sigma^2, \sigma_u^2, \beta_1, u_{1:(T_0+t-1)}, B_{1:(T_0+t-1)} | y_{1:(T_0+t)}). \tag{A12}$$

We used 10,000 iterations to estimate each of the components of the predictive likelihood.