Appendix

A1. Proofs of Theorems in §3.3

To prove Theorem 1, we introduce a new objective function for the dynamic programming problem defined in Eq. (1): minimizing the time to complete all projects in $\mathcal{S}$; for simplicity, we call this objective the time objective and the objective in Eq. (1) the GDP objective. We define $T(\mathcal{X})$ to be the minimum time to complete all projects in $\mathcal{S} \setminus \mathcal{X}$ given that $\mathcal{X}$ is already completed. We have the following DP formulation for the time objective,

$$\begin{aligned} T(\mathcal{X}) &= \min_{i \in \mathcal{S} \setminus \mathcal{X}} \{ \frac{c_i}{rg(\mathcal{X})} + T(\mathcal{X} \cup \{i\}) \}, \\ T(\mathcal{S}) &= 0. \end{aligned} \tag{29}$$

Note that $T(\mathcal{X})$ differs from $\tilde{T}(\mathcal{X})$. The following theorem characterizes the optimal project sequence for the DP with the time objective (Eq. 29), see the Extended Appendix for a proof.

Theorem 4. *Consider the projects in $\mathcal{S}$ with $v_1 \leq v_2 \leq \ldots \leq v_S$; under the conditions of constant or increasing c_j/v_j (linear or diseconomies of scale), the optimal project sequence is $1, 2, ..., S-1, S$ for the dynamic programming problem defined in Eq. (29).*

Theorem 4 implies that it is optimal to build smaller projects first to minimize the time of completing all projects under the conditions of linear and diseconomies of scale.

We establish the following lemma (see the Extended Appendix for a proof).

Lemma 6.1. *The following mathematical properties holds*

1. $\beta x \geq 1 - e^{-\beta x}$ *for* $x \geq 0$.
2. *Let* $h(x, b) = \frac{xe^{-\beta x}}{1 - be^{-\beta x}}$ *for* $x \geq 0$, $b \geq 1$ *and* $\beta > 0$, *then* $\lim_{x \to 0} h(x, b) = \frac{1}{b\beta}$, $\lim_{x \to \infty} h(x, b) = 0$, *and* $h(x, b)$ *is non-increasing in* x *for any* b.

Proof of Theorem 1: First, we consider $\mathcal{S}$ with two projects (1 and 2) and $v_1 \leq v_2$. Let $\mathcal{X}$ be a sequenced set and define $\tilde{V}(\mathcal{X})$ to be the total discounted GDP contributed by $\mathcal{X}$ by building it up following its sequence. By Eq. (1), we have

$$\begin{aligned} V(\emptyset) &= \max\{\tilde{V}(\{1,2\}), \tilde{V}(\{2,1\})\}, \\ \tilde{V}(\{1,2\}) &= \frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} + e^{-\beta \frac{c_1}{rg_0}} \frac{av_2}{\beta} e^{-\beta \frac{c_2}{r(g_0 + av_1)}}, \\ \tilde{V}(\{2,1\}) &= \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} + e^{-\beta \frac{c_2}{rg_0}} \frac{av_1}{\beta} e^{-\beta \frac{c_1}{r(g_0 + av_2)}}. \end{aligned}$$

Hence,

$$\begin{aligned} \tilde{V}(\{1,2\}) - \tilde{V}(\{2,1\}) &= A - B, \\ A &= \frac{av_1}{\beta} (e^{-\beta \frac{c_1}{rg_0}} - e^{-\beta \frac{c_2}{rg_0}} e^{-\beta \frac{c_1}{r(g_0 + av_2)}}), \\ B &= \frac{av_2}{\beta} (e^{-\beta \frac{c_2}{rg_0}} - e^{-\beta \frac{c_1}{rg_0}} e^{-\beta \frac{c_2}{r(g_0 + av_1)}}). \end{aligned}$$

Clearly, A and B are positive; they can be re-written as,

$$\begin{aligned} A &= \frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} e^{\beta \frac{ac_1 v_2}{rg_0(g_0+av_2)}}), \\ B &= \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} e^{\beta \frac{ac_2 v_1}{rg_0(g_0+av_1)}}). \end{aligned}$$

Let $b_1 = e^{\beta \frac{ac_1 v_2}{rg_0(g_0+av_2)}}$ and $b_2 = e^{\beta \frac{ac_2 v_1}{rg_0(g_0+av_1)}}$. Clearly $b_2 \geq b_1 \geq 1$ because $v_1 \leq v_2$ and $c_1/v_1 \leq c_2/v_2$. Hence,

$$\begin{aligned} B/A &= \frac{v_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_2)}{v_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)} \leq \frac{v_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_1)}{v_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)}, \\ &\leq \frac{c_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_1)}{c_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)} = h(\frac{c_2}{rg_0}, b_1)/h(\frac{c_1}{rg_0}, b_1) \leq 1. \end{aligned}$$

The first inequality holds because $b_1 \leq b_2$, the second inequality holds because $c_1/v_1 \leq c_2/v_2$, and the last inequality follows by $c_1 \leq c_2$ and Lemma 6.1.

Then we consider the general case of $\mathcal{S}$ with projects $1, 2, ..., S$ and $v_1 \leq v_2 \leq \ldots \leq v_S$. We make the induction assumption that Theorem 1 holds for any subset of $\mathcal{S}$. By Eq. (1),

$$V(\emptyset) = \max\{\frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} + V(\{1\}), \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} + V(\{2\}), \ldots, \frac{av_S}{\beta} e^{-\beta \frac{c_S}{rg_0}} + V(\{S\})\}. \tag{30}$$

By the induction assumption, the optimal sequence corresponding to $V(\{1\})$ must be $2, ..., S-1, S$, and thus the sequenced set for the first term in the right-hand-side of Eq. (30) is $\{1, 2, ..., S-1, S\}$. Similarly, the sequenced set for the second term of the right-hand-side of Eq. (30) is $\{2, 1, 3, ..., S-1, S\}$, and so on; finally, the sequenced set for the last (Sth) term of the right-hand-side of Eq. (35) (see Extended Appendix I) is $\{S, 1, 2, ..., S-1\}$.

By the induction assumption, $\tilde{V}(\{1, 2, 3, ..., S-1\}) \geq \tilde{V}(\{2, 1, 3, ..., S-1\})$. In addition,

$$\begin{aligned} \tilde{V}(\{1, 2, 3, ..., S-1, S\}) &= \tilde{V}(\{1, 2, 3, ..., S-1\}) + \frac{av_S}{\beta} e^{-\beta(\tilde{T}(\{1,2,3,\ldots,S-1\}) + \frac{c_S}{rg(\{1,2,3,\ldots,S-1\})})}, \\ \tilde{V}(\{2, 1, 3, ..., S-1, S\}) &= \tilde{V}(\{2, 1, 3, ..., S-1\}) + \frac{av_S}{\beta} e^{-\beta(\tilde{T}(\{2,1,3,\ldots,S-1\}) + \frac{c_S}{rg(\{2,1,3,\ldots,S-1\})})}. \end{aligned}$$

Because $\tilde{T}(\{1, 2, 3, ..., S-1\}) \leq \tilde{T}(\{2, 1, 3, ..., S-1\})$ (by Theorem 4) and $g(\{1, 2, 3, ..., S-1\}) = g(\{2, 1, 3, ..., S-1\})$, then we must have $\tilde{V}(\{1, 2, 3, ..., S-1, S\}) \geq \tilde{V}(\{2, 1, 3, ..., S-1, S\})$ (that is, adding an identical project to the end of two sequenced sets does not change the order of their total discounted GDPs). The same logic can be applied to the third through $S-1$th terms of the right-hand-side of Eq. (30) to get similar results.

For the last term in the right-hand-side of Eq. (35), we first observe that $\tilde{V}(\{S, 1, 2, ..., S-1\}) \leq \tilde{V}(\{1, S, 2, ..., S-1\})$ because $\tilde{V}(\{S, 1\}) \leq \tilde{V}(\{1, S\})$ (by induction assumption) and the following sequence of $2, ..., S-1$ is identical for both sets. We further observe that $\tilde{V}(\{1, S, 2, ..., S-1\}) \leq \tilde{V}(\{1, 2, ..., S-1, S\})$ because the first project in both sequenced sets is identical, and $\tilde{V}(\{S, 2, ..., S-1\}) \leq$

$\tilde{V}(\{2, ..., S-1, S\})$ (by induction assumption). Combining these observations, we have $\tilde{V}(\{S, 1, 2, ..., S-1\}) \le \tilde{V}(\{1, 2, ..., S-1, S\})$. The rest of the proof is straightforward. □

Proof of Theorem 2:

To see if the sequence of building smaller projects first is also optimal under economies of scale, we study the case of $\mathcal{S}$ with only two projects (1 and 2), and $v_1 \le v_2$, $c_1 \le c_2$ but $c_1/v_1 > c_2/v_2$. By Eq. (1), we have

$$\begin{aligned} V(\emptyset) &= \max\{\tilde{V}(\{1,2\}), \tilde{V}(\{2,1\})\}, \\ \tilde{V}(\{1,2\}) &= \frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} + e^{-\beta \frac{c_1}{rg_0}} \frac{av_2}{\beta} e^{-\beta \frac{c_2}{r(g_0+av_1)}}, \\ \tilde{V}(\{2,1\}) &= \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} + e^{-\beta \frac{c_2}{rg_0}} \frac{av_1}{\beta} e^{-\beta \frac{c_1}{r(g_0+av_2)}}, \end{aligned}$$

where $\tilde{V}(\{1,2\})$ ($\tilde{V}(\{2,1\})$) is the total discounted GDP contributed by the sequenced set $\{1,2\}$ ($\{2,1\}$, respectively). Hence,

$$\tilde{V}(\{1,2\}) - \tilde{V}(\{2,1\}) = \frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1) - \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_2),$$

where $b_1 = e^{\beta \frac{ac_1 v_2}{rg_0(g_0+av_2)}}$ and $b_2 = e^{\beta \frac{ac_2 v_1}{rg_0(g_0+av_1)}}$. Note that when $\frac{g_0}{a} >> 1$, we will have $b_1 > b_2 \ge 1$ because of $c_1 v_2 > c_2 v_1$ (economies of scale). Let $A = \frac{av_1}{\beta} e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)$, $B = \frac{av_2}{\beta} e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_2)$, and $\mu_1 = v_1/c_1$ and $\mu_2 = v_2/c_2$. Then in these cases,

$$\begin{aligned} B/A &= \frac{v_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_2)}{v_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)} \ge \frac{v_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_1)}{v_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)}, \\ &= \frac{\mu_2}{\mu_1} \frac{c_2 e^{-\beta \frac{c_2}{rg_0}} (1 - e^{-\beta \frac{c_1}{rg_0}} b_1)}{c_1 e^{-\beta \frac{c_1}{rg_0}} (1 - e^{-\beta \frac{c_2}{rg_0}} b_1)} = \frac{\mu_2}{\mu_1} h(\frac{c_2}{rg_0}, b_1)/h(\frac{c_1}{rg_0}, b_1). \end{aligned}$$

By $c_1/v_1 > c_2/v_2$, we must have $\frac{\mu_2}{\mu_1} > 1$. Fixing project parameters but let $\frac{1}{rg_0} \to 0$, $h(\frac{c_2}{rg_0}, b_1)/h(\frac{c_1}{rg_0}, b_1) \to 1$ (Lemma 6.1). Thus, there must exist a threshold such that for all rg_0 greater than this threshold, $B/A > 1$, which implies that in these cases, the optimal sequence is to build the larger project first, and thus the sequence of building smaller projects first is not optimal. □

Proof of Theorem 3: It is easy to see that the second term in Eq. (2) is linear in x, so we just need to prove that the first term in this equation is convex in x. To this end, we note that

$$\tilde{\omega}^{1-x} = \tilde{\omega}\tilde{\omega}^{-x} = \tilde{\omega} e^{-x ln(\tilde{\omega})},$$

where $-ln(\tilde{\omega}) \ge 0$ because $\tilde{\omega}$ is a fraction. It is easy to see that for any $b \ge 0$, e^{bx} and xe^{bx} are convex in x. Thus, the first term in Eq. (2) is convex in x. □

A2. Model Parameters for Pakistan

We first note that the hydro plant setup cost, IC_j, includes cost of transmission line and grid station that connect the hydro plant to the nearest transmission system. It also includes the cost of water

Hydro Plant:		
$A_{j,t}$	Net external inflow to segment j of the river system in period t	Unit: Billion Cubic Meters
T_j	Setup time for the hydro plant at location j	Unit: Period
IC_j	Setup cost for the hydro plant at location j	Unit: Million US $
OC_j	Annual operating cost of the hydro plant at location j	Unit: Million US $
PC_j	Installed capacity (electricity) of the hydro plant at location j	Unit: MW
CR_j	Conversion ratio of the hydro plant at location j	Unit: MW / Billion Cubic Meters
C_j	The storage capacity of the hydro plant at location j	Unit: Billion Cubic Meters
$X_{j,j'}$	Matrix of immediate upstream locations adjusted by river yield loss	$\%$
Power Supply:		
ω	% yield in power transmission per 100 miles	95.5%
Ω_{jk}	Transmission yield in % between location j and demand zone k	$\Omega_{jk} = \omega^{D_{jk}}$
R_k^E	Regional power demand ration	Unit: %
Power Demand:		
$RP_{t'}$	Electricity price in year t'	Unit:Million US $
$OEC_{t'}$	Power consumption from sources other than hydro in year t'	Unit:MW
Water Supply:		
γ	% yield in irrigation distribution per 100 miles	55%
Γ_{jk}	Water distribution yield in % between location j and demand zone k	$\Gamma_{jk} = \gamma^{D_{jk}}$
η	% yield by evaporation for stored water in reservoirs per period	80%
r_j	% of total flow into location j that can be withdrawn for usage	31%
G_{kt}^W	Water gap at demand zone $k \in \mathcal{K}_W$ in period t	Unit: Billion Cubic Meters
$Y_{j,k}$	Water distribution feasibility matrix	Binary
R_k^W	Regional Water demand ratio	Unit: %
Water Demand:		
$HE_{t'}$	Arable land in year t'	Unit: Hectare
DR^W	Water demand ratio of wet season	Unit: %
DR^D	Water demand ratio of dry season	Unit: %
EW	Existing water supply for agriculture from sources other than hydro	Unit: Billion Cubic Meters
Flood:		
MT_j	Peak tolerance flow in one period at location j	Unit: Billion Cubic Meter
ξ_j	Fraction of flooded water running away from the river at location j	$\%$
Miscellaneous:		
$RA_{t'}$	Ration in year t'. i.e. % of GDP	Unit: %
$\beta_{t'}$	Discount factor in year t'	Unit: %
D_{jk}	Distance between location j and demand zone k	Unit: 100 *miles*
Z_k	The maximum share of the total consumption (power and water) for demand zone k	Unit: %
M	A real number large enough for the Big-M method	$200,000$

Table 4: Model parameters for Pakistan's hydro supply chain.

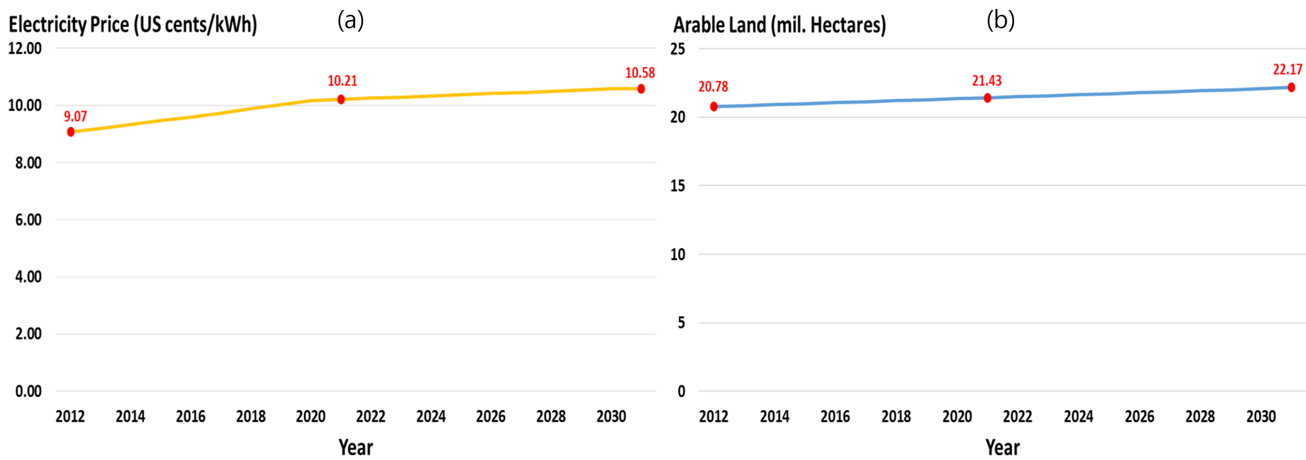

Figure 8: The projections of electricity price (a) and arable land (b).

passages (or pumping station) that connect the plant to the nearest canal system. All other parameters of hydro plants are extracted from feasibility studies and government research reports. Based on the alternating current (AC) technology and voltages used in Pakistan, we estimate the transmission yield loss per 100 miles, $(1-\omega)$, to be 4.5% (Hurlbut 2012). The energy and water demands are estimated by econometric models such as, linear regression, non-stationarity (augmented Dickey-Fuller) tests and cointegration models (Studenmund 2011, see Appendix A3 for details). Energy consumption met by sources other than hydro is estimated based on the government projection of the power consumption and the projected energy mix (see Appendix A3 for details).

$X_{j,j'}$ is calculated based on river seepage losses in Pakistan (Saeed and Khan 2014) and the distance between locations j' and j. r_j is determined by the general statistics of water withdrawn at hydro plant locations (Barber 2009). The water distribution feasibility matrix, $Y_{j,k}$, is determined by Assumption 1 where $Y_{j,k} = 0$ if the water demand zone k is upstream of location j, $Y_{j,k} = 1$ otherwise. The MT_j can be estimated by the average flow per period at location j plus a location dependent buffer (from government's statistics). By Tariq and van de Giesen (2012), we set $\xi_j = 1$ for lower Indus river locations in Sindh and Balochistan, and $\xi_j = 0$ elsewhere. Z_k is set to be the percentage of contribution made by each demand zone in either power or water.

A3. Empirical Study on Pakistan

Here we summarize data collection and processing, related literature and variable selection, key testing results and regression equations for the empirical study on Pakistan. We also provide a model validation. We leave the detailed tests and estimation processes to the Extended Appendix.

The literature shows a strong dependence of GDP on capital, labor, and energy consumption (Tang and Shahbaz 2013), water consumption (Rosegrant, et al. 2000) and flood (ADB 2013). Inspired by the literature and given the available data, we shall study the impact of these variables on GDP (agriculture, non-agriculture, total) for Pakistan. We obtained data on capital, labor force, total GDP and agriculture GDP (from 1961 to 2011) for Pakistan from the World Bank. We also obtained water withdraw data (total and agriculture, from 1961 to 2011), electricity price (1970-2011) and arable land (1961-2011) from the World Bank and Pakistan government but with some missing years. We obtained complete canal water withdraw data from WAPDA but it is not the total agriculture water withdraw (which also includes ground water). To handle the missing data, we use simple interest fit (i.e., linear extrapolation, also used by the World Bank) on the ratio between the total and canal water withdraw (due to the potential correlation between them) and then calculate the total water withdraw from the canal water withdraw. The total energy consumption data (1961-2011) of Pakistan is obtained from the World Bank. However, the data on agriculture energy consumption is available only from 1990 to 2011. The flood related data (1977-2011) is available from Pakistan Federal Flood Commission. Although we have complete data on flooded area, flood volume data is missing for some years. To handle missing data, we first calculate the average height of the flooded water by the total flood volume over total flood areas over the years with flood and data available, then we estimate the flood volume by multiplying the average height of water with the flooded area.

We first study the impact of the estimated flood volume on flood economic loss (real loss, GDP inflator adjusted) from 1977 to 2011 and find a significant correlation. Specifically, the real loss and estimated flood volume are stationary time series and strongly cointegrated (at a 5% level), thus we can run a linear regression model between them (e.g., Studenmund 2011) which is significant ($R^2 = 85\%$ and p-value $\approx 10^{-15}$). We also study the impact of the estimated flood volume on agriculture GDP (1977-2011), and find that despite their cointegration, there isn't a significant correlation between them ($R^2 = 10\%$, p-value $= 0.07$).

Ideally, we need data on agriculture GDP before flood damage is accounted for. However, there is no breakdown on flood loss between agriculture and non-agriculture sectors, and so we have to use agriculture GDP after flood damage is accounted for as an approximation. Fortunately, this approximation is reasonable because the impact of flood on agriculture GDP is insignificant as is verified above. We calculate non-agriculture GDP by subtracting agriculture GDP and adding flood loss to the GDP. We find that agriculture GDP depends strongly on factors such as water consumption, agricultural capital, and agricultural labor for Pakistan. To remove the multicollinearity issue, we normalize the data by agriculture labor and focus on the dependence of agricultural GDP per agricultural labor on agricultural capital per agricultural labor and the agricultural water consumption per agricultural labor. We obtain data from the World Bank for 1992-2010 on these time series but there is no data on water consumption, we have to convert the water withdraw data to consumption data by the national average distribution loss of 40% (Saeed and Khan 2014). We find that all time series are non-stationary but strongly cointegrated and the residuals are not autocorrelated. We also find that the independent variables are not strongly correlated (the absolute value of correlation coefficient < 0.7 and variance inflation factor < 10) and thus there is no multi-collinearity issue. The regression model has a R^2=92.9%, the coefficients of the two independent variables are significant (p-value ≈ 0.00), and no autocorrelation issue exists (Durbin Watson statistic $= 1.83$).

For non-agriculture GDP, we find that it depends strongly on energy consumption, non-agricultural capital and non-agricultural labor. To remove multicollinearity issue, we focus on non-agriculture GDP per non-agricultural labor, energy consumption per non-agricultural labor and non-agricultural capital per non-agricultural labor. From the World Bank, we obtained the data for 1993-2007. We find that all these time series are non-stationary but strongly cointegrated. The independent variables are not strongly correlated and thus there is no multi-collinearity issue. The linear regression model is significant with $R^2 = 89.7\%$ and p-value≈ 0.00 for all independent variables.

To validate the GDP models, we first train the models (following procedures listed above) using data up to 2007 for Pakistan. We then use the models to predict GDP for the years from 2008 to 2011 with the actual capital, labor, water and energy consumption, and flood data in these years. We compare the predicted GDP with the actual GDP for the four years and find that average percentage difference between predicted and actual GDPs ((actual GDP - predicted GDP)/actual GDP) is 4.02% with the minimum 2.6% (2009) and the maximum 5.99% (2010). In 2010, Pakistan had the worst flood for the last 50 years.

Inspired by the literature which shows that the electricity demand strongly depends on previous year's GDP and current year's electricity price (see, e.g., Zaman, et al. 2015), we obtained the GDP and electricity consumption data (1970 to 2011) from the World Bank and the electricity price data from National Transmission and Despatch Company of Pakistan. We find that all variables are non-stationary but strongly cointegrated; the independent variables are not strongly correlated and thus there is no multi-collinearity issue. The linear regression model, described in Eq. (25), is significant with $R^2 = 98.9\%$ and p-value ≈ 0.00 for previous year's GDP (the coefficient $b_1 = 0.1$) and p-value < 0.011 for current year's electricity price (the coefficient $b_2 = 179.9$).

Hussain, et al. (2011) shows that water demand for irrigation depends strongly on previous year's agriculture GDP and arable land. We obtain data (1989-2003) from the World Bank and find that all variables are non-stationary but strongly cointegrated. The independent variables are not strongly correlated and thus there is no multi-collinearity issue. The linear regression model, described in Eq. (26), is significant with $R^2 = 95.8\%$ and p-value ≈ 0.00 for previous year's GDP ($b_3 = 0.00148$) and p-value ≈ 0.003 for arable land ($b_4 = 0.0000037$).

To project electricity prices, we use the government's projections (NTDC 2011). To predict arable land for the next 20 years, we compute the average growth rate for 1961 to 2011 by logarithmic trendline (Figure 8). We project the power consumption met by other resources than hydro in year $t^{'}$, OEC_t, based on the government projection of the power consumption (NTDC 2014) for the peak total demand in the period of 2011-2036, and the projected energy mix (Perwez and Sohail 2014).

Capital in agricultural and non-agricultural sectors may be dependent on the previous year's GDP (Uneze 2013). We obtain data (1962-2011) from the World Bank and find that both dependent and independent variables in the capital models (agricultural and non-agricultural) are non-stationary but strongly cointegrated. The linear regression model of non-agricultural capital, see Eq. (31), is significant with $R^2 = 99.6\%$ and p-value ≈ 0.00 for previous year's non-agricultural GDP. The linear regression model of agricultural capital, see Eq. (32), is significant with $R^2 = 99.2\%$ and p-value ≈ 0.00 for previous year's agricultural GDP.

$$k_{t'}^{NA} = k_{t'-1}^{NA} + 4.44 \cdot (g_{t'-1}^{NA} - g_{t'-2}^{NA}) \qquad \text{for} \quad t' = 1, \ldots, T/2. \tag{31}$$

$$k_{t'}^{A} = k_{t'-1}^{A} + 5.55 \cdot (g_{t'-1}^{A} - g_{t'-2}^{A}) \qquad \text{for} \quad t' = 1, \ldots, T/2. \tag{32}$$

Labor force in agricultural and non-agricultural sectors may also depend on the previous year's GDP (Okun 1962). We obtain data (1975-2000) from the World Bank and find that both dependent and independent variables in the labor models (agricultural and non-agricultural) are non-stationary but strongly cointegrated. The linear regression model of non-agricultural labor force, see Eq. (33), is significant with $R^2 = 98.9\%$ and p-value ≈ 0.00 for previous year's non-agricultural GDP. The linear regression model of agricultural labor force, Eq. (34), is significant with $R^2 = 95.3\%$ and p-value ≈ 0.00 for previous year's agricultural GDP.

$$l_{t'}^{NA} = l_{t'-1}^{NA} + 0.00045 \cdot (g_{t'-1}^{NA} - g_{t'-2}^{NA}) \qquad \text{for} \quad t' = 1, \ldots, T/2. \tag{33}$$

$$l_{t'}^{A} = l_{t'-1}^{A} + 0.0016 \cdot (g_{t'-1}^{A} - g_{t'-2}^{A}) \qquad \text{for} \quad t' = 1, \ldots, T/2. \tag{34}$$