A Decision-Analytic Framework to explore the water-energy-food nexus in complex and transboundary water resources systems, with Climate Change Uncertainty

Koundouri, Phoebe and Englezos, Nikos and Kartala, Xanthi and Tsionas, Mike

January 2019

ABSTRACT

In this paper we develop and apply a stochastic multistage dynamic cooperative game for managing transboundary water resources, within the water-food-energy nexus framework, under climate uncertainty. The mathematical model is solved for the non-cooperative and cooperative (Stackelberg "*leader–follower*") cases and is applied to the Omo-Turkana River Basin in Africa. The empirical application of the model calls for sector-specific production function estimations, for which we employ nonparametric treatment of the production functions a la Gandhi, Navarro, and Rivers (2017), while we extend it to allow for technical inefficiency in production and autocorrelated TFP. Bayesian analysis is performed using a Sequential Monte Carlo / Particle-Filtering approach. We find that the cooperative solution is the optimal pathway not only for both riparian countries, but for the sustainable use of the basin as well, whereas in extreme Climate Change circumstances it remains the welfare maximizing option. We argue that the detail and sophistication of both the mathematical and econometric models are needed for robust policy recommendations towards sustainable management of transboundary resources.

Key words: stochasticity, Markov processes, endogenous adaptation, technical inefficiency, autocorrelation, copula approach.

1. THE ECONOMIC MODEL

There is a substantial body of literature on stochastic water resource management from which only few studies exist on the influence of stochastic water resource management on transboundary water sharing. *Bhaduri et al.* (2011) investigated the uncertainty in water resource management in a transboundary water sharing problem and evaluated the scope and sustainability for a potential cooperative agreement between countries. On the other hand, *Kim et al.* (1989) studied a deterministic renewable groundwater optimal management problem in the face of two-sector linear demands, while *Koundouri and Christou* (2006) revisited this problem under the presence of a backstop technology.

Bhaduri et al. (2011) utilized a stochastic differential Stackelberg "*leader–follower*" game to produce qualitative results on the optimal transboundary water allocation between an upstream and a downstream area. In their model, the upstream area represents the *leader* and applies his strategy first, *a priori* knowing that the *follower* downstream area observes its actions and *a posteriori* moves accordingly. In view of *Kim et al.* (1989) and *Koundouri and Christou* (2006), this paper extends the stochastic game problem of *Bhaduri et al.* (2011) to capture quantitatively the influence of stochastic water resources on transboundary water allocation over multiple sectors of the economy, following now a multistage dynamic cooperative game framework even in the case of climate change. A main contribution of this framework is the introduction of the five core water economic sectors taking into account their dependence with the social benefit of water use per country. Furthermore, in contrast to *Bhaduri et al.* (2011) who had to restrict the leader's strategy space to quadratic functions of the state variable in order to obtain a sub-optimal qualitative solution of the problem, we maximize the leader's objective function, using the follower's reaction strategy, over all possible strategies to provide an optimal solution of our stochastic game problem that is also quantitatively tractable.

We assume that water resources evolve through time and follow a geometric Brownian motion. However, the characteristics of Brownian motion in terms of variance are different between the upstream and the downstream country, based on the assumption that the effects of climate change are regionally different. Additionally, we are able to determine how the water abstraction of the riparian countries will change in the long run, taking into account the greater variability of water availability caused by climate change. In other words, the suggested model describes water allocation between the upstream and the downstream country in such a case, with and without any cooperation in water sharing, taking into account how uncertainty in water supply affects the water abstraction rates of the countries, and explores the underlying conditions that may influence decisions on water allocations.

The upstream country has the upper riparian right to unilaterally divert water while the freshwater availability of the downstream one partially depends on the water usage in the upstream country. We denote the countries by superscripts, where *U*denotes the upstream country and *D*stands for the downstream country.

Following *Bhaduri et al*. (2011), we consider at first a complete filtered probability space $(\Omega, \Im, \Im_t, P)$ and we assume that water flow is stochastic and uncertainty in the flow of water can be attributed to climate change. Then the annual renewable water resource due to the river basin, *W*., evolves through time according to the Geometric Brownian motion:

$$dW_t = \sigma^W W_t dz_t^W \quad , \quad t \geq 0, \tag{1}$$

where $z_{\cdot}^W$ is a standard Wiener process and σ^W can be considered as the volatility of water flow in the upstream country.

Let us denote by $w_{i\cdot}^h$ the total freshwater utilization and by T_i^h the **exit time** of the *i*-th sector, per country *h = U, D,*and for each sector*i*= 1, 2, …, 5, together with the convention of $T_0 \equiv 0$ and $T_5 \equiv \infty$. Then the change in the level of water resources available in the upstream country, $W_{\cdot}^U$, for the *j*-th exit stage is represented by

$$dW_{jt}^U = \left[W_t - \sum_{i=j}^{5} w_{it}^U \right] \text{dt}, \quad T_{j-1}^U \leq t < T_j^U, \quad j = 1, 2, ..., 5. \tag{2}$$

The water availability in the downstream country depends on the total water consumption in the upstream one and runoff, denoted by *R*., which is also stochastic in the model. Thus, the latter can be expressed through another Geometric Brownian motion as

$$dR_t = \sigma^R R_t dz_t^R, \quad t \geq 0, \tag{3}$$

where $z_{\cdot}^R$ is another standard Wiener process independent of $z_{\cdot}^W$. Thus, the stock of water in the downstream area, where agricultural products and fisheries are produced, is denoted by $S_{.}$, and is a function of the stochastic water resources and the control variables $w_{\cdot}^{h} = (w_{1\cdot}^{h}, w_{2\cdot}^{h}, \ldots, w_{5\cdot}^{h})$ per country *h = U, D*; in fact, for the *(j,k)*-th exit stage of the upstream and downstream countries, respectively, it follows the dynamics:

$$dS_{jkt} = \left\{ W_t - \sum_{i=j}^{5} w_{it}^{U} - \sum_{l=k}^{5} w_{lt}^{D} + R_t - O_t \right\} dt, \quad T_{j-1}^{U} \leq t < T_j^{U} \text{ and } T_{k-1}^{D} \leq t < T_k^{D}, \quad j,k = 1,2,\ldots,5, \tag{4}$$

where $S(0)=S_0$ is an initial condition. Here, O_t denotes the outflow and evaporation of water from this area and can be formulated by a third Geometric Brownian motion as

$$dO_t = \sigma^{O} O_t dz_t^{O}, \quad t \geq 0, \tag{5}$$

where $z_{\cdot}^{O}$ is a third standard Wiener process independent of $z_{\cdot}^{W}$ and $z_{\cdot}^{R}$.

We assume that the inverse demand function for the water utilization of the *j*-th exit stage, per sector *i*and country *h*, is represented by

$$p_{jt}^{h} = \frac{a_i^h}{b_i^h} - \frac{1}{b_i^h} \cdot w_{it}^{h}, \quad T_{j-1}^{h} \leq t < T_j^{h}, \quad i = j,,\ldots,5, \quad j = 1,2,\ldots,5, \quad h = U, D, \tag{6}$$

where $p_{j\cdot}^{h}$ is the price of water at each stage *j*, which is the same for the different sectors, and $a_i^h \in \mathbb{R}$, $b_i^h > 0$ are constant sector-specific demand parameters. The sector-specific inverse demand curves are ordered so that $a_1^h / b_1^h < a_2^h / b_2^h < \cdots < a_5^h / b_5^h$, which implies that water demand for each of the five sectors reaches zero sequentially over time as the price of water increases over time, leading to the endogenously defined exit times T_j^h, $j = 1,2,\ldots,5,$ of the five economic sectors per country *h = U, D.* Here, aggregate water demand turns out to be a piecewise linear function.

Since consumers are the only group deriving benefits from water, the inverse demand curve is the marginal social benefit curve. Hence, consider further the benefit of water consumption w_i^h per sector *i* of country*h*, namely social benefit (*SB*), as

$$SB_i^h\left(w_i^h\right) = \int \left(\frac{a_i^h}{b_i^h} - \frac{1}{b_i^h} \cdot w_i^h \right) dw_i^h = \frac{a_i^h}{b_i^h} w_i^h - \frac{1}{2b_i^h} \cdot (w_i^h)^2 + c_i^h, \quad h = U, D, \tag{7}$$

where c_i^h is a constant that corresponds to other factors of production (variables) such as labour, capital, natural capital, etc. (cf. Section 3). It is obvious that the benefit function is strictly concave for all possible values of w_i^h.

Water abstraction from rivers may be taken directly from the flowing waters in the channel (surface water abstraction) or can be achieved through inter-basin flow transfer schemes. Thus, we may assume that the marginal extraction cost (MC) for the *j*-th exit stage of the upstream country is a decreasing function of the available water resources W^U of the form:

$$MC^U(W_j^U) = k_2^U - k_1^U W_j^U, \quad j = 1,2,\ldots,5,$$

where k_1^U, $k_2^U > 0$ are given constants. In fact, we consider that as water becomes increasingly scarce in the economy, the government would exploit water through appropriating and purchasing a greater share of aggregate economic output, in terms of dams, pumping stations, supply infrastructure, etc. (*Barbier*, 2000). Given the high cost of building infrastructure and expanding supplies, this will lead to a higher marginal cost of water. Then the total cost (*TC*) function of water withdrawing w_i^U from the river per sector *i* =*j*, …, 5,for the *j*-th exit stage of the upstream country is given by

$$TC^U\left(W_j^U, w_i^U\right) = (k_2^U - k_1^U W_j^U) w_i^U, \quad i = j,...,5, \quad j = 1,2,...,5, \tag{8}$$

which is an increasing function of the water extraction variable. On the other hand, the downstream country extracts water from its available stock, thus for the (*j,k*)-th exit stage the *MC* of the downstream country is a decreasing function of the available water stock S_{jk} and has the form:

$$MC^D(S_{jk}) = k_2^D - k_1^D S_{jk}, \quad j,k = 1,2,...,5,$$

where k_1^D, $k_2^D > 0$ are given constants. Then the *TC* function of water withdrawing w_l^D from the water stock in the downstream country per sector *l* = *k*, …, 5for the (*j,k*)-th exit stage is given by

$$TC^D\left(S_{jk}, w_l^D\right) = (k_2^D - k_1^D S_{jk}) w_l^D, \quad l = k,...,5, \tag{9}$$

which is an increasing function of the water extraction variables.

Finally, the downstream country receives benefits from storing water, as the net consumer surplus or economic benefit from food (agricultural product and fisheries) production, denoted by the concave quadratic function of water stock S_{jk} per (*j,k*)-th exit stage:

$$F^D(S_{jk}) = \eta_1^D S_{jk}^2, + \eta_2^D S_{jk} + \eta_3^D, \quad j,k = 1,2,...,5, \tag{10}$$

where $\eta_1^D < 0$, $\eta_2^D \in \mathbb{R}$, $\eta_3^D \in \mathbb{R}$ are constants.

1.1 Non-Cooperative Approach

We present below a non-cooperative framework, where there is no any agreement between the two countries regarding either water or food sharing.

Upstream Case

The upstream country chooses the economically potential rate of water utilization that maximizes its own net benefit (*NB*)per *j*-th exit stage, which can be expressed as

$$NB_j^U = \sum_{i=j}^{5} SB_i^U(w_i^U) - \sum_{i=j}^{5} TC^U\left(W_j^U, w_i^U\right). \tag{10}$$

Thus, the upstream country maximization problem can be formulated as follows:

$$
\begin{aligned}
J^U &= \max_{w^U_{\cdot}} \sum_{j=1}^{5} J_j^U = \max_{w^U_{\cdot}} \sum_{j=1}^{5} E\left\{\int_{T_{j-1}^U}^{T_j^U} e^{-rt} NB_j^U \, dt\right\} \\
&= \max_{w^U_{\cdot}} \sum_{j=1}^{5} E\left\{\int_{T_{j-1}^U}^{T_j^U} e^{-rt} \sum_{i=j}^{5}\left[SB_i^U(w_{it}^U) - TC^U\left(W_{jt}^U, w_{it}^U\right)\right]dt\right\} \qquad (11)\\
&= \max_{w^U_{\cdot}} \sum_{j=1}^{5} E\left\{\int_{T_{j-1}^U}^{T_j^U} e^{-rt} \sum_{i=j}^{5}\left[\frac{a_i^U}{b_i^U} w_{it}^U - \frac{1}{2b_i^U}\cdot (w_{it}^U)^2 + c_i^U - (k_2^U - k_1^U W_{jt}^U) w_{it}^U\right]dt\right\},
\end{aligned}
$$

where J_j^U stands for the upstream country's net social benefit of the *j*-th exit stage, *j*=1,2,…,5, and $w^U_{\cdot} = (w^U_{1\cdot},\ w^U_{2\cdot},\ \ldots,\ w^U_{5\cdot})$ is the sectorial water extraction vector process for the upstream country, subject to the river basin annual renewable water resource equation of (1) and the upstream country water resources (state) equation of(3).

For the*j*-exit stage we have the Hamiltonian:

$$
H_j^U\left(W_{jt}^U, w_{jt}^U, \ldots, w_{5t}^U, \lambda_{jt}^U\right) \triangleq \sum_{i=j}^{5}\left[\frac{a_i^U}{b_i^U} w_{it}^U - \frac{1}{2b_i^U}\cdot (w_{it}^U)^2 + c_i^U - (k_2^U - k_1^U W_{jt}^U) w_{it}^U\right] + \lambda_{jt}^U\left[W_t - \sum_{i=j}^{5} w_{it}^U\right], \quad j = 1, 2, \ldots, 5,
$$

where $\lambda^U_{j\cdot}$ is the *j*-exit stage adjoint variable that represents water scarcity rents for the upstream country. The necessary conditions for optimality are given as follows:

$$
\frac{\partial H_j^U}{\partial w_{it}^U} = \frac{a_i^U}{b_i^U} - \frac{1}{b_i^U}\cdot w_{it}^U - (k_2^U - k_1^U W_{jt}^U) - \lambda_{jt}^U = 0, \quad i = j, \ldots, 5, \qquad (12)
$$

$$
d\lambda_{jt}^U = \left[-\frac{\partial H_j^U}{\partial W_{jt}^U} + r\lambda_{jt}^U\right]dt \iff d\lambda_{jt}^U = \left[-k_1^U \sum_{i=j}^{5} w_{it}^U + r\lambda_{jt}^U\right]dt. \qquad (13)
$$

From the first optimality condition, we have:

$$
w_{it}^U = a_i^U - b_i^U (k_2^U - k_1^U W_{jt}^U) - b_i^U \lambda_{jt}^U, \quad i = j, \ldots, 5.
$$

Then substituting to the state equation we have:

$$
dW_{jt}^U = \left[-k_1^U\left(\sum_{i=j}^{5} b_i^U\right) W_{jt}^U + \left(\sum_{i=j}^{5} b_i^U\right)\lambda_{jt}^U + W_t - \sum_{i=j}^{5} a_i^U + k_2^U \sum_{i=j}^{5} b_i^U\right]dt
$$

while substituting to the adjoint equation we have

$$
d\lambda_{jt}^U = \left\{-\left(k_1^U\right)^2\left(\sum_{i=j}^{5} b_i^U\right) W_{jt}^U + \left[k_1^U\left(\sum_{i=j}^{5} b_i^U\right) + r\right]\lambda_{jt}^U - k_1^U \sum_{i=j}^{5} a_i^U + k_1^U k_2^U \sum_{i=j}^{5} b_i^U\right\}dt.
$$

Setting $A_j^U \triangleq \sum_{i=j}^{5} a_i^U$ and $B_j^U \triangleq \sum_{i=j}^{5} b_i^U$ we obtain the forward-backward differential equations system (FBDEs):

$$dW_{jt}^{U}=\left[-k_1^U B_j^U W_{jt}^U + B_j^U \lambda_{jt}^U + W_t - A_j^U + k_2^U B_j^U\right]dt,$$
$$d\lambda_{jt}^{U}=\left[-\left(k_1^U\right)^2 B_j^U W_{jt}^U + \left(k_1^U B_j^U + r\right)\lambda_{jt}^U - k_1^U A_j^U + k_1^U k_2^U B_j^U\right]dt,$$
$$W_0^U = w_0, \quad \lim_{t\to\infty}\lambda_{jt}^U = 0, \quad j=1,2,...,5.$$

To solve the above system of FBDEs we impose a solution of the form:

$$\lambda_{jt}^U = N_{jt}^U W_{jt}^U + M_{jt}^U, \quad i = j,...,5,$$

where N_{jt}^U and M_{jt}^U are functions to be determined. Taking differentials we have:

$$\begin{aligned} d\lambda_{jt}^U &= W_{jt}^U dN_{jt}^U + N_{jt}^U dW_{jt}^U + dM_{jt}^U \\ &= W_{jt}^U dN_{jt}^U + dM_{jt}^U + \left\{-k_1^U B_j^U W_{jt}^U N_{jt}^U + B_j^U W_{jt}^U \left(N_{jt}^U\right)^2 + B_j^U N_{jt}^U M_{jt}^U + \left[w_t A_j^U + k_2^U B_j^U\right] N_{jt}^U\right\} dt \end{aligned}$$

while from the backward equation of the system we have:

$$d\lambda_{jt}^{U}=\left[-\left(k_1^U\right)^2 B_j^U W_{jt}^U + \left(k_1^U B_j^U + r\right) W_{jt}^U N_{jt}^U + \left(k_1^U B_j^U + r\right) M_{jt}^U - k_1^U A_j^U + k_1^U k_2^U B_j^U\right]dt.$$

A sufficient condition for the latter to be equal

$$dN_{jt}^U = \left[-B_j^U \left(N_j^U\right)^2 + \left(2k_1^U B_j^U + r\right) N_{jt}^U - \left(k_1^U\right)^2 B_j^U\right]dt,$$
$$\lim_{t\to\infty} N_{jt}^U = 0,$$

which is a backward Riccatti equation (BRE) that can be solved numerically.

And

$$dM_{jt}^U = \left[\left(-B_j^U N_j^U + k_1^U B_j^U + r\right) M_{jt}^U - \left(w_t - A_j^U + k_2^U B_j^U\right) N_j^U - k_1^U A_j^U + k_1^U k_2^U B_j^U\right]dt,$$
$$\lim_{t\to\infty} M_{jt}^U = 0,$$

which given the solution of Riccatti for N_{j}^U is a backward linear first-order ordinary differential equation (ODE).

Substituting the linear solution form to the forward equation of the FBSDEs system, we have:

$$\begin{aligned} &dW_{jt}^U = \left[\left(-k_1^U + N_{jt}^U\right) B_{jt}^U W_{jt}^U + B_j^U M_j^U + w_t - A_j^U + k_2^U B_j^U\right]dt, \\ &W_{00}^U = w_0, \end{aligned} \tag{14}$$

which is a forward linear ODE. Then the backward adjoint variable follows from the linear transformation and the optimal water use follows from the optimality condition.

Downstream Case

On the other hand, the downstream country water consumption/production depends on the inflow from the upstream country, and the runoff generated within the country's share of the water stock in the downstream area. Based on the given availability of water, the downstream country maximizes its NBper exit stage (j,k) quantified as follows:

$$NB_{jk}^{D}=\sum_{l=k}^{5}SB_{l}^{D}\left(w_{l}^{D}\right)+F^{D}(S_{jk})-\sum_{l=k}^{5}TC^{D}\left(S_{jk},w_{l}^{D}\right). \qquad (15)$$

Hence, the downstream country maximization problem is given by

$$\begin{aligned}
J^{D}&=\max_{w_{t}^{D}}\sum_{k=1}^{5}\sum_{j=1}^{5}J_{jk}^{D}=\max_{w_{t}^{D}}\sum_{k=1}^{5}\sum_{j=1}^{5}E\left\{\int_{\left\{T_{j-1}^{U}\le t<T_{j}^{U}\right\}\cap\left\{T_{k-1}^{D}\le t<T_{k}^{D}\right\}}e^{-rt}NB_{jk}^{D}dt\right\}\\
&=\max_{w_{t}^{D}}\sum_{k=1}^{5}\sum_{j=1}^{5}E\left\{\int_{\left\{T_{j-1}^{U}\le t<T_{j}^{U}\right\}\cap\left\{T_{k-1}^{D}\le t<T_{k}^{D}\right\}}e^{-rt}\begin{bmatrix}\sum_{l=k}^{5}SB_{l}^{D}\left(w_{lt}^{D}\right)+F^{D}(S_{jkt})\\-\sum_{l=k}^{5}TC^{D}\left(S_{jkt},w_{lt}^{D}\right)\end{bmatrix}dt\right\}\\
&=\max_{w_{t}^{D}}\sum_{k=1}^{5}\sum_{j=1}^{5}E\left\{\int_{\left\{T_{j-1}^{U}\le t<T_{j}^{U}\right\}\cap\left\{T_{k-1}^{D}\le t<T_{k}^{D}\right\}}e^{-rt}\begin{bmatrix}\sum_{l=k}^{5}\left(\frac{a_{l}^{D}}{b_{l}^{D}}w_{lt}^{D}-\frac{1}{2b_{l}^{D}}\cdot(w_{lt}^{D})^{2}+c_{l}^{D}\right)+\eta_{1}^{D}S_{jkt}^{2}\\+\eta_{2}^{D}S_{jkt}+\eta_{3}^{D}-(k_{2}^{D}-k_{1}^{D}S_{jkt})\sum_{l=k}^{5}w_{lt}^{D}\end{bmatrix}dt\right\},
\end{aligned}$$

(16)

where J_{jk}^{D} represents the downstream country's net social benefit of the (*j,k*)-thexit stage, *j*, *k*=1,2,…,5, and $w_{t}^{D}=(w_{1t}^{D},\ w_{1t}^{D},\ ...,\ w_{5t}^{D})$ is the sectorial water extraction vector process for the downstream country, subject to the river basin annual renewable water resource equation of (1), the upstream country water resources equation of (2), the runoff flow equation of (3), the outflow equation of (5), and the stock of water (state variable) in the downstream area equation of (4).

For the(*j,k*)-th exit-stage we have the Hamiltonian:

$$\begin{aligned}
H_{jk}^{D}\left(S_{jk},w_{kt}^{D},...,w_{5t}^{D},\lambda_{jkt}^{D}\right)=&\sum_{l=k}^{5}\left[\frac{a_{l}^{D}}{b_{l}^{D}}w_{lt}^{D}-\frac{1}{2b_{l}^{D}}\cdot(w_{lt}^{D})^{2}+c_{l}^{D}\right]+\eta_{1}^{D}S_{jkt}^{2}\\
&+\eta_{2}^{D}S_{jkt}+\eta_{3}^{D}-(k_{2}^{D}-k_{1}^{D}S_{jkt})\sum_{l=k}^{5}w_{lt}^{D}\\
&+\lambda_{jkt}^{D}\left[W_{t}-\sum_{i=j}^{5}w_{it}^{U}-\sum_{l=k}^{5}w_{lt}^{D}+R_{t}-O_{t}\right],\qquad j,k=1,2,...,5,
\end{aligned}$$

where λ_{jkt}^{D} is the (*j,k*)-thexit stage adjoint variable that represents water scarcity rents for the downstream country. The necessary conditions for optimality are given as follows:

$$\frac{\partial H_{jk}^{D}}{\partial w_{lt}^{D}}=\frac{a_{l}^{D}}{b_{l}^{D}}-\frac{1}{b_{l}^{D}}\cdot w_{lt}^{D}-(k_{2}^{D}-k_{1}^{D}S_{jkt})-\lambda_{jkt}^{D}=0,\quad l=k,...,5, \qquad (17)$$

$$d\lambda_{jkt}^{D}=\left[-\frac{\partial H_{jk}^{D}}{\partial S_{jkt}}+r\lambda_{jkt}^{D}\right]dt \quad \Leftrightarrow \quad d\lambda_{jkt}^{D}=\left[-k_1^{D}\sum_{l=k}^{5}w_{lt}^{D}-2\eta_1^{D}S_{jkt}-\eta_2^{D}+r\lambda_{jkt}^{D}\right]dt. \quad (18)$$

From (18) we have that

$$\lambda_{jkt}^{D}=\frac{a_l^{D}}{b_l^{D}}-\frac{1}{b_l^{D}}\cdot w_{lt}^{D}-(k_2^{D}-k_1^{D}S_{jkt}),\ \ l=k,...,5,\ \ (20)$$

which implies that

$$d\lambda_{jkt}^{D}=-\frac{1}{b_l^{D}}\cdot dw_{lt}^{D}+k_1^{D}\left[W_t-\sum_{i=j}^{5}w_{it}^{U}-\sum_{l=k}^{5}w_{lt}^{D}+R_t-O_t\right]dt.$$

Substituting the last two relationships back to (19) we obtain:

$$-\frac{1}{b_l^{D}}\cdot dw_{lt}^{D}+k_1^{D}\left[W_t-\sum_{i=j}^{5}w_{it}^{U}-\sum_{l=k}^{5}w_{lt}^{D}+R_t-O_t\right]dt$$

$$=\left\{-k_1^{D}\sum_{l=k}^{5}w_{lt}^{D}-2\eta_1^{D}S_{jkt}-\eta_2^{D}+r\left[\frac{a_l^{D}}{b_l^{D}}-\frac{1}{b_l^{D}}\cdot w_{lt}^{D}-(k_2^{D}-k_1^{D}S_{jkt})\right]\right\}dt$$

which simplifies to

$$dw_{lt}^{D}=\left\{\begin{array}{l} r\left[w_{lt}^{D}+b_l^{D}(k_2^{D}-k_1^{D}S_{jkt})-a_l^{D}\right]+2b_l^{D}\eta_1^{D}S_{jkt} \\ +b_l^{D}\eta_2^{D}+b_l^{D}k_1^{D}\left[W_t-\sum_{i=j}^{5}w_{it}^{U}+R_t-O_t\right] \end{array}\right\}dt$$

for $l = k, \ldots, 5$.

From the first optimality condition, we have that:

$$w_{lt}^{D}=a_l^{D}-b_l^{D}(k_2^{D}-k_1^{D}S_{jkt})-b_l^{D}\lambda_{jkt}^{D}, \quad l=k,...,5.$$

Then substituting to the state equation we have:

$$dS_{jkt}=\left[-k_1^{D}\left(\sum_{l=k}^{5}b_l^{D}\right)S_{jkt}+\left(\sum_{l=k}^{5}b_l^{D}\right)\lambda_{jkt}^{D}+W_t-\sum_{i=j}^{5}w_{it}^{U}+R_t-O_t-\sum_{l=k}^{5}a_l^{D}+k_2^{D}\sum_{l=k}^{5}b_i^{D}\right]dt$$

while substituting to the adjoint equation we have

$$d\lambda_{jkt}^{D}=\left\{\left[-\left(k_1^{D}\right)^2\left(\sum_{l=k}^{5}b_l^{D}\right)-2\eta_1^{D}\right]S_{jkt}+\left[k_1^{D}\left(\sum_{l=k}^{5}b_l^{D}\right)+r\right]\lambda_{jkt}^{D}-k_1^{D}\sum_{l=k}^{5}a_l^{D}+k_1^{D}k_2^{D}\sum_{l=k}^{5}b_l^{D}-\eta_2^{D}\right\}dt$$

Setting $A_j^{U} \triangleq \sum_{i=j}^{5}a_i^{U}$ and $B_j^{U} \triangleq \sum_{i=j}^{5}b_i^{U}$ we obtain the system of FBDEs:

$$dS_{jkt} = \left[-k_1^D B_k^D S_{jkt} + B_k^D \lambda_{jkt}^D + W_t - \sum_{i=j}^{5} w_{it}^U + R_t - O_t - A_k^D + k_2^D B_k^D \right] dt,$$

$$d\lambda_{jkt}^D = \left\{ \left[-\left(k_1^D\right)^2 B_k^D - 2\eta_1^D \right] S_{jkt} + \left[k_1^D B_k^D + r \right] \lambda_{jkt}^D - k_1^D A_k^D + k_1^D k_2^D B_k^D - \eta_2^D \right\} dt,$$

$$S_{000} = s_0, \quad \lim_{t\to\infty} \lambda_{jkt}^D = 0, \quad j,k = 1,2,\ldots,5.$$

To solve the above system of FBDEs we impose a solution of the form:

$$\lambda_{jkt}^D = N_{jkt}^D S_{jkt} + M_{jkt}^D,$$

where N_{jk}^D and M_{jk}^D are functions to be determined. Taking differentials, we have:

$$\begin{aligned} d\lambda_{jkt}^D &= S_{jkt} dN_{jkt}^D + N_{jkt}^D dS_{jkt} + dM_{jkt}^D \\ &= S_{jkt} dN_{jkt}^D + dM_{jkt}^D \\ &+ \left\{ -k_1^D B_k^D S_{jkt} N_{jkt}^D + B_k^D S_{jkt} \left(N_{jkt}^D\right)^2 + B_k^D N_{jkt}^D M_{jkt}^D + \left[W_t - \sum_{i=j}^{5} w_{it}^U + R_t - O_t - A_k^D + k_2^D B_k^D \right] N_{jkt}^D \right\} dt, \end{aligned}$$

while from the backward equation of the system we have:

$$d\lambda_{jkt}^D = \left\{ \left[-\left(k_1^D\right)^2 B_k^D - 2\eta_1^D \right] S_{jkt} + \left[k_1^D B_k^D + r \right] S_{jkt} N_{jkt}^D + \left[k_1^D B_k^D + r \right] M_{jkt}^D - k_1^D A_k^D + k_1^D k_2^D B_k^D - \eta_2^D \right\} dt.$$

A sufficient condition for the latter to be equal is given by

$$dN_{jkt}^D = \left[-B_k^D \left(N_{jkt}^D\right)^2 + \left(2k_1^D B_k^D + r\right) N_{jkt}^D - \left(k_1^D\right)^2 B_k^D - 2\eta_1^D \right] dt,$$

$$\lim_{t\to\infty} N_{jkt}^D = 0,$$

which is a BRE that can be solved numerically.

Also

$$dM_{jkt}^D = \begin{bmatrix} \left(-B_k^D N_{jkt}^D + k_1^D B_k^D + r\right) M_{jkt}^D - \left(W_t - \sum_{i=j}^{5} w_{it}^U + R_t - O_t - A_k^D + k_2^D B_k^D \right) N_{jkt}^D \\ - k_1^D A_k^D + k_1^D k_2^D B_k^D - \eta_2^D \end{bmatrix} dt,$$

$$\lim_{t\to\infty} M_{jkt}^D = 0,$$

which given the solution of Riccatti for N_{jk}^D is a backward linear first-order ordinary differential equation (ODE).

Substituting the linear solution form to the forward equation of the FBSDEs system, we have:

$$dS_{jkt} = \left[\left(-k_1^D + N_{jkt}^D\right) S_{jkt} + B_k^D M_{jkt}^D + W_t - \sum_{i=j}^{5} w_{it}^U + R_t - O_t - A_k^D + k_2^D B_k^D \right] dt, \quad (19)$$

$$S_{000} = s_0,$$

which is a forward linear ODE. Then the backward adjoint variable follows from the linear transformation and the optimal water use follows from the optimality condition.

1.2 COOPERATIVE APPROACH

The 2016 power sharing agreement provides a mandate for the Kenya-Ethiopia Electricity Highway Project (or the Eastern Electricity Highway Project), which will let the construction of a 1,000km power line to run from Ethiopia to Kenya to be completed by 2018.[1] The agreement is built upon an MoU signed in 2006 between the Ethiopian Electric Power Corporation and the Kenya Electricity Transmission Company for the joint development of the project.[2] The environmental and social impact assessment report was approved in 2012, although it has been criticised as it was conducted after any objection could be made.[3] Following a World Bank loan of US$684 million,[4] construction began in June 2016.[5] While the 2016 agreement is not yet publicly available, it is reported that the agreement will allow Ethiopia to supply Kenya with 400 megawatts of hydro-power at less than 1 US cent/kwh.[6] However, the hydro-power source (or sources) that will supply this transmission line is not officially stated, although the World Bank modified an official project report specifying that power would be sourced "from Ethiopia's Gilgel Gibe hydropower scheme",[7] changing the reference to the dam in its next report instead to "Ethiopia's power grid".[8]

In this section, we present the model of the inter-sectoral water sharing strategy between the upstream and downstream country in a cooperative setting. In particular, the downstream country offers a discounted price for food exports to the upstream country, in exchange for greater transboundary water flow that results in a higher water reserve accumulation and sequentially in a higher production of food. In what follows, we utilize a differential Stackelberg "*leader–follower*" game to determine the inter-sector optimal water allocation between the upstream and downstream country. The conditions for stability in water sharing are explored with respect to increasing variance in water flow due to climate change. The upstream country represents the *leader* and applies his strategy first, *a priori* knowing that the *follower* downstream country, observes its actions and *a posteriori* moves accordingly. First, we find the solution to the follower's problem of maximizing a payoff function, and then, using the follower's reaction strategy, we maximize the leader's objective function.

Since all the model coefficients are deterministic functions of time, we assume that the respective countries use *Markovian perfect strategies*. These strategies are decision rules that dictate optimal action of the respective players, conditional on the current values of the state variables (upstream level of water resources, level of water stock reserves downstream, etc), that summarize the latest available information of the dynamic system. The Markovian perfect strategies determine a *subgame perfect equilibrium* and define an *equilibrium set of decisions* dependent on previous actions.

---

[1]("Ethiopia, Kenya to enhance cooperation on energy sector.," n.d.)
[2]("Kenya-Ethiopia Electricity Highway, Kenya," n.d.)
[3](Abbink, 2012)
[4]("AFCC2/RI-The Eastern Electricity Highway Project under the First Phase of the Eastern Africa Power Integration Program," n.d.)
[5]("Kenya-Ethiopia Electricity Highway, Kenya," n.d.)
[6]("Ethiopia, Kenya to enhance cooperation on energy sector.," n.d.)
[7](*Resettlement Action Plan (RAP) Final Report 2012*, n.d.)
[8](*Resettlement Policy Framework Draft Report 2012*, n.d.)

Downstream Case

Given the intersectoral water abstraction policy $w_{\cdot}^{U} = \left(w_{1\cdot}^{U}, \ldots, w_{5\cdot}^{U}\right)$ that is announced by the upstream country, the downstream country is faced with the same maximization problem with the one of the non-cooperative case, i.e., maximize (16)subject to the state equations (2) - (6). For every *j,k*=1,2,…,5, the (*j,k*)-th exit stage Hamiltonian of the system is also given by (17) and its necessary conditions for optimality by (18) and(19). From (18) the optimal response policy of the downstream country is represented by

$$w_{lt}^{D} = a_l^{D} - b_l^{D}(k_2^{D} - k_1^{D} S_{jkt}) - b_l^{D} \lambda_{jkt}^{D}, \quad l = k, \ldots, 5, \quad j = 1, \ldots, 5, \tag{20}$$

which together with(18) we have that

$$\lambda_{jkt}^{D} = \frac{a_l^{D}}{b_l^{D}} - \frac{1}{b_l^{D}} \cdot w_{lt}^{D} - (k_2^{D} - k_1^{D} S_{jkt}), \quad l = k, \ldots, 5, \quad j = 1, \ldots, 5, \tag{21}$$

which implies that

$$d\lambda_{jkt}^{D} = -\frac{1}{b_l^{D}} \cdot dw_{lt}^{D} + k_1^{D} \left[W_t - \sum_{i=j}^{5} w_{it}^{U} - \sum_{l=k}^{5} w_{lt}^{D} + R_t - O_t \right] dt.$$

Substituting the last two relationships back to (20) we obtain:

$$\begin{aligned} & -\frac{1}{b_l^{D}} \cdot dw_{lt}^{D} + k_1^{D} \left[W_t - \sum_{i=j}^{5} w_{it}^{U} - \sum_{l=k}^{5} w_{lt}^{D} + R_t - O_t \right] dt \\ & = \left\{ -k_1^{D} \sum_{l=k}^{5} w_{lt}^{D} - 2\eta_1^{D} S_{jkt} - \eta_2^{D} + r \left[\frac{a_l^{D}}{b_l^{D}} - \frac{1}{b_l^{D}} \cdot w_{lt}^{D} - (k_2^{D} - k_1^{D} S_{jkt}) \right] \right\} dt, \end{aligned}$$

which simplifies to

$$dw_{lt}^{D} = \left\{ \begin{aligned} & r \left[w_{lt}^{D} + b_l^{D}(k_2^{D} - k_1^{D} S_{jkt}) - a_l^{D} \right] + 2 b_l^{D} \eta_1^{D} S_{jkt} \\ & + b_l^{D} \eta_2^{D} + b_l^{D} k_1^{D} \left[W_t - \sum_{i=j}^{5} w_{it}^{U} + R_t - O_t \right] \end{aligned} \right\} dt, \quad l = k, \ldots, 5. \tag{22}$$

The second optimality condition may be written

$$d\lambda_{jkt}^{D} = \left\{ -k_1^{D} \sum_{l=k}^{5} \left[a_l^{D} - b_l^{D}(k_2^{D} - k_1^{D} S_{jkt}) - b_l^{D} \lambda_{jkt}^{D} \right] - 2\eta_1^{D} S_{jkt} - \eta_2^{D} + r \lambda_{jkt}^{D} \right\} dt,$$

which implies to

$$d\lambda_{jkt}^{D} = \left\{ -k_1^{D} \sum_{l=k}^{5} a_l^{D} - \sum_{l=k}^{5} b_l^{D}(k_2^{D} - k_1^{D} S_{jkt}) - 2\eta_1^{D} S_{jkt} - \eta_2^{D} + \left(r + k_1^{D} \sum_{l=k}^{5} b_l^{D} \right) \lambda_{jkt}^{D} \right\} dt. \tag{23}$$

Also, the stock of the water state equation can be written as

$$dS_{jkt} = \left[W_t - \sum_{i=j}^{5} w_{it}^{U} + R_t - O_t - \sum_{l=k}^{5} a_l^{D} + \sum_{l=k}^{5} b_l^{D}(k_2^{D} - k_1^{D} S_{jkt}) + \sum_{l=k}^{5} b_l^{D} \lambda_{jkt}^{D} \right] dt. \tag{24}$$

Upstream Case

The upstream country receives now food benefits from the downstream country denoted by the concave quadratic function of water stock S_{jk} per (*j,k*)-th exit stage:

$$F^U(S_{jk}) = \eta_1^U S_{jk}^2, + \eta_2^U S_{jk} + \eta_3^U, \qquad j,k = 1,2,\ldots,5,$$

where $\eta_1^U < 0,\ \eta_2^U \in \mathbb{R},\ \eta_3^U \in \mathbb{R}$ are constants, and its *NB* function is given by

$$NB_{jk}^U = \sum_{i=j}^{5} SB_i^U(w_{it}^U) + F^U\left(S_{jkt}\right) - \sum_{i=j}^{5} TC^U\left(W_{jt}^U, w_{it}^U\right), \qquad j,k = 1,2,\ldots,5. \tag{25}$$

Therefore, the upstream country, anticipating the downstream country's optimal response as analysed in the previous case, chooses the optimal water abstraction vector process $w_{\cdot}^U = (w_{1\cdot}^U,\ w_{2\cdot}^U,\ \ldots,\ w_{5\cdot}^U)$ under cooperation by solving the following maximization problem:

$$\begin{aligned}
J^U &= \max_{w_{\cdot}^U} \sum_{j=1}^{5}\sum_{k=1}^{5} J_{jk}^U = \max_{w_{\cdot}^U} \sum_{j=1}^{5}\sum_{k=1}^{5} E\left\{ \int\limits_{\{T_{j-1}^U \le t \le T_j^U\} \cap \{T_{k-1}^D \le t \le T_k^D\}} e^{-rt} NB_{jk}^U dt \right\} \\
&= \max_{w_{\cdot}^U} \sum_{j=1}^{5}\sum_{k=1}^{5} E\left\{ \int\limits_{\{T_{j-1}^U \le t \le T_j^U\} \cap \{T_{k-1}^D \le t \le T_k^D\}} e^{-rt} \left[\begin{array}{l} \sum_{i=j}^{5} SB_i^U(w_{it}^U) + F_{jk}^U\left(S_{jkt}\right) \\ -\sum_{i=j}^{5} TC^U\left(W_{jt}^U, w_{it}^U\right) \end{array} \right] dt \right\} \\
&= \max_{w_{\cdot}^U} \sum_{j=1}^{5}\sum_{k=1}^{5} E\left\{ \int\limits_{\{T_{j-1}^U \le t \le T_j^U\} \cap \{T_{k-1}^D \le t \le T_k^D\}} e^{-rt} \left[\begin{array}{l} \sum_{i=j}^{5}\left(\frac{a_i^U}{b_i^U} w_{it}^U - \frac{1}{2b_i^U}\cdot (w_{it}^U)^2 + c_i^U \right) \\ + \eta_1^U S_{jkt}^2, + \eta_2^U S_{jkt} + \eta_3^U - (k_2^U - k_1^U W_{jt}^U)\sum_{i=j}^{5} w_{it}^U \end{array} \right] dt \right\},
\end{aligned} \tag{26}$$

subject to the state equation of the (*j,k*)-th exit stage (3) and Hamiltonian system of the downstream country, i.e. (25) and (26).

For the (*j,k*)-th exit stage we have the augmented Hamiltonian:

$$\begin{aligned}
H_{jk}^U\left(W_{jt}^U, S_{jkt}, \lambda_{jkt}^D, w_{jt}^U, \ldots, w_{5t}^U, \mu_{j\kappa t}, \nu_{j\kappa t}, \xi_{j\kappa t}\right) &= \sum_{i=j}^{5}\left[\frac{a_i^U}{b_i^U} w_{it}^U - \frac{1}{2b_i^U}\cdot (w_{it}^U)^2 + c_i^U - (k_2^U - k_1^U W_{jt}^U) w_{it}^U \right] \\
&+ \eta_1^U S_{jkt}^2 + \eta_2^U S_{jkt} + \eta_3^U + \mu_{jkt}\left[W_t - \sum_{i=j}^{5} w_{it}^U \right] \\
&+ \nu_{jkt}\left[W_t - \sum_{i=j}^{5} w_{it}^U + R_t - O_t - \sum_{l=k}^{5} a_l^D + \sum_{l=k}^{5} b_l^D (k_2^D - k_1^D S_{jkt}) + \sum_{l=k}^{5} b_l^D \lambda_{jkt}^D \right] \\
&+ \xi_{jkt}\left\{ -k_1^D\left[\sum_{l=k}^{5} a_l^D - \sum_{l=k}^{5} b_l^D (k_2^D - k_1^D S_{jkt}) \right] - 2\eta_1^D S_{jkt} - \eta_2^D + \left(r + k_1^D \sum_{l=k}^{5} b_l^D \right)\lambda_{jkt}^D \right\},
\end{aligned}$$

where $\left(\mu_{j\kappa\cdot}, \nu_{j\kappa\cdot}, \xi_{j\kappa\cdot}\right)$ is the vector of the associated adjoint variables. The necessary conditions for optimality are given below:

$$\frac{\partial H_{jk}^U}{\partial w_{it}^U} = \frac{a_i^U}{b_i^U} - \frac{1}{b_i^U}\cdot w_{it}^U - (k_2^U - k_1^U W_{jt}^U) - \mu_{jkt} - \nu_{jkt} = 0, \quad i = j,...,5, \tag{27}$$

$$d\mu_{jkt} = \left[-\frac{\partial H_{j\kappa}^U}{\partial W_{jt}^U} + r\mu_{jkt} \right] dt \quad \Leftrightarrow \quad d\mu_{jkt} = \left[-k_1^U \sum_{i=j}^{5} w_{it}^U + r\mu_{jkt} \right] dt, \tag{28}$$

$$\begin{gathered} dv_{jkt} = \left[-\frac{\partial H_{j\kappa}^U}{\partial S_{jkt}} + r\nu_{jkt} \right] dt \quad \Leftrightarrow \\ dv_{jkt} = \left\{ -2\eta_1^U S_{jkt} - \eta_2^U + k_1^D \left(\sum_{l=k}^{5} b_l^D \right) \nu_{jkt} + \left[\left(k_1^D \right)^2 \sum_{l=k}^{5} b_l^D + 2\eta_1^D \right] \xi_{jkt} + r\nu_{jkt} \right\} dt, \end{gathered} \tag{29}$$

$$d\xi_{jkt} = \left[-\frac{\partial H_{j\kappa}^U}{\partial \lambda_{jt}^D} + r\xi_{jkt} \right] dt \quad \Leftrightarrow \quad d\xi_{jkt} = \left[-\sum_{l=k}^{5} b_l^D \nu_{jkt} - k_1^D \sum_{l=k}^{5} b_l^D \xi_{jkt} \right] dt, \qquad \xi(0) = 0. \tag{30}$$

From the first optimality condition we have

$$w_{it}^U = a_i^U - b_i^U (k_2^U - k_1^U W_{jt}^U) - b_i^U \mu_{jkt} - b_i^U \nu_{jkt}, \quad i = j,...,5,$$

which implies to

$$\mu_{jkt} + \nu_{jkt} = \frac{a_i^U}{b_i^U} - \frac{1}{b_i^U} w_{it}^U - (k_2^U - k_1^U W_{jt}^U), \qquad i = j,...,5, \tag{31}$$

and

$$d\left(\mu_{jkt} + \nu_{jkt} \right) = -\frac{1}{b_i^U} dw_{it}^U + k_1^U \left[W_t - \sum_{i=j}^{5} w_{it}^U \right] dt.$$

Adding by parts (28), (29) and (31) and substituting in the resulting equation the last two relationships, we get that

$$-\frac{1}{b_i^U}\cdot dw_{it}^U + k_1^U \left[W_t - \sum_{i=j}^{5} w_{it}^U \right] dt$$

$$= \left\{ \begin{aligned} & -k_1^U \sum_{i=j}^{5} w_{it}^U - 2\eta_1^U S_{jkt} - \eta_2^U + k_1^D \sum_{l=k}^{5} b_l^D \nu_{jkt} + \left[\left(k_1^D \right)^2 \sum_{l=k}^{5} b_l^D + 2\eta_1^D \right] \xi_{jkt} \\ & + r\left[\frac{a_l^U}{b_l^U} - \frac{1}{b_l^U}\cdot w_{it}^U - (k_2^U - k_1^U \, \mathrm{W}_{jt}^U) \right] \end{aligned} \right\} dt,$$

which simplifies to

$$dw_{it}^U = \left\{ \begin{aligned} & r\left[w_{it}^U + b_i^U (k_2^U - k_1^U \, \mathrm{W}_{jt}^U) - a_i^U \right] + k_1^U b_i^U W_t + 2\eta_1^U b_i^U S_{jkt} + \eta_2^U b_i^U \\ & - k_1^D b_i^U \left(\sum_{l=k}^{5} b_l^D \right) \nu_{jkt} - b_i^U \left[\left(k_1^D \right)^2 \sum_{l=k}^{5} b_l^D + 2\eta_1^D \right] \xi_{jkt} \end{aligned} \right\} dt, \qquad i = j,...,5. \tag{32}$$

The relationships in (32) forms a system of differential equations to be solved for the optimal water abstraction $w_{i\cdot}^{U}$ per each sector *i* =1, 2, …, 5 by the upstream country, in the leader-follower cooperative setting explained above, and the respective endogenous switching times T_j^U , *j* =1, 2, …, 5.

From the first optimality condition we have:

$$w_{it}^{U} = a_i^U - b_i^U (k_2^U - k_1^U W_{jt}^U) - b_i^U \mu_{jkt} - b_i^U \nu_{jkt}, \quad i = j, ..., 5. \tag{33}$$

Then the state equation may be written as

$$dW_{jt}^{U} = \left[W_t - A_j^U + B_j^U (k_2^U - k_1^U \, \mathrm{W}_{jt}^U) + B_j^U \mu_{jkt} + B_j^U \nu_{jkt} \right]$$

which implies that

$$\begin{aligned} dW_{jt}^{U} &= \left[-k_1^U B_j^U W_{jt}^U + B_j^U \mu_{jkt} + B_j^U \nu_{jkt} + W_t - A_j^U + k_2^U B_j^U \right] dt, \\ W_{00}^{U} &= w_0^U . \end{aligned} \tag{34}$$

Similarly,

$$\begin{aligned} dS_{jkt} &= \begin{bmatrix} -k_1^U B_j^U W_{jt}^U - k_1^D B_k^D S_{jkt} + B_j^D \lambda_{jkt} + B_j^U \mu_{jkt} + B_j^U \nu_{jkt} + W_t - A_j^U + k_2^U B_j^U \\ + R_t + O_t - A_k^D + k_2^D B_k^D \end{bmatrix} dt, \\ S_{000} &= s_0, \end{aligned} \tag{35}$$

and

$$\begin{aligned} d\lambda_{jkt}^{D} &= \left\{ - \left[\left(k_1^D \right)^2 B_k^D + 2\eta_1^D \right] S_{jkt} + \left(r + k_1^D B_k^D \right) \lambda_{jkt}^D - k_1^D A_k^D + k_1^D k_2^D B_k^D - \eta_2^D \right\} dt, \\ \lim_{t\to\infty} \lambda_{jkt}^{D} &= 0. \end{aligned} \tag{36}$$

Furthermore, the adjoint equation may take the equivalent form

$$d\mu_{jkt} = \left[- \left(k_1^U \right)^2 B_j^U W_{jt}^U + \left(r + k_1^U B_j^U \right) \mu_{jkt} + k_1^U B_j^U \nu_{jkt} - k_1^U A_j^U + k_1^U k_2^U B_j^U \right] dt,$$

$$\lim_{t\to\infty} \mu_{jkt} = 0.$$

$$d\nu_{jkt} = \left\{ -2\eta_1^U S_{jkt} + \left(r + k_1^D B_k^D \right) \nu_{jkt} + \left[\left(k_1^D \right)^2 B_k^D + 2\eta_1^D \right] \xi_{jkt} - \eta_2^U \right\} dt,$$

$$\lim_{t\to\infty} \nu_{jkt} = 0.$$

$$d\xi_{jkt} = \left[-B_k^D \nu_{jkt} - k_1^D B_k^D \xi_{jkt} \right] dt,$$

$$\xi_{jk0} = 0.$$

Assume that $\eta_1^D = \eta_2^D = \eta_3^D = 0$ and $\eta_1^U = 0$. We are looking for solutions of the form

$$\nu_{jkt} = N_{jkt} \xi_{jkt} + M_{jkt} \tag{37}$$

where $N_{jk\cdot}$ and $M_{jk\cdot}$ are functions to be determined.

Taking differentials, we have

$$d\nu_{jkt} = N_{jkt} d\xi_{jkt} + \xi_{jkt} dN_{jkt} + dM_{jkt} = \xi_{jkt} dN_{jkt} + dM_{jkt} + \left(-B_k^D N_{jkt}\nu_{jkt} - k_1^D B_k^D N_{jkt}\xi_{jkt}\right)dt,$$

while the relationship for $\nu_{jk\cdot}$ may be written equivalently

$$d\nu_{jkt} = \left\{\left[\left(r + k_1^D B_k^D\right)N_{jkt} + \left(k_1^D\right)^2 B_k^D\right]\xi_{jkt} + \left(r + k_1^D B_k^D\right)M_{jkt} - \eta_2^U\right\}dt.$$

A sufficient condition for the latter to be equivalent to

$$dN_{jkt} = \left[-B_k^D\left(N_{jk}^U\right)^2 + \left(2k_1^D B_k^D + r\right)N_{jkt}^U + \left(k_1^D\right)^2 B_k^D\right]dt,$$
$$\lim_{t\to\infty} N_{jkt} = 0$$

which given the solution of the Riccati for $N_{jk\cdot}$ is a backward linear first order ODE.

Substituting the linear solution form to the forward equation for $\xi_{jk\cdot}$ we have

$$d\xi_{jkt} = \left[-B_k^D\left[N_{jkt} + k_1^D\right]\xi_{jkt} - B_k^D M_{jkt}\right]dt,$$
$$\xi_{jk0} = 0,$$

which is a forward linear ODE. Then the backward variable $\nu_{jk\cdot}$ follows from the linear transformation.

Given that $\nu_{jk\cdot}$ and $\xi_{jk\cdot}$ are known to be the previous scheme. impose the following linear transformation between the variables:

$$\mu_{jkt} = \Lambda_{jkt} W_{jt}^U + \Xi_{jkt}$$

where $\Lambda_{jk\cdot}$ and $\Xi_{jk\cdot}$ are functions to be determined.

Taking differentials, we have

$$\begin{aligned} d\mu_{jkt} &= \Lambda_{jkt} dW_{jt}^U + W_{jt}^U d\Lambda_{jkt} + d\Xi_{jkt} \\ &= W_{jt}^U d\Lambda_{jkt} + d\Xi_{jkt} \\ &\quad + \left\{\left[B_j^U \Lambda_{jkt}^2 - k_1^U B_j^U \Lambda_{jkt}\right]W_{jt}^U + B_j^U \Lambda_{jkt}\Xi_{jkt} + B_j^U \nu_{jkt}\Lambda_{jkt} + \left[W_t - A_j^U + k_2^U B_j^U\right]\Lambda_{jkt}\right\}dt \end{aligned}$$

while the relationship for $\mu_{jk\cdot}$ may be formulated as

$$d\mu_{jkt} = \left\{\left[-\left(k_1^U\right)^2 B_j^U + \left(k_1^U B_j^U + r\right)\Lambda_{jkt}\right]W_{jt}^U + \left(k_1^U B_j^U + r\right)\Xi_{jkt} + k_1^U B_j^U \nu_{jkt} - k_1^U A_j^U + k_1^U k_2^U B_j^U\right\}dt$$
.

A sufficient condition for the latter to be equivalent is

$$d\Lambda_{jkt} = \left[-B_j^U \Lambda_{jkt}^2 + \left(2k_1^U B_j^U + r\right)\Lambda_{jkt} - \left(k_1^U\right)^2 B_j^U\right]dt,$$
$$\lim_{t\to\infty} \Lambda_{jkt} = 0,$$

which is a backward Riccati that can be solved numerically. And

$$d\Xi_{jkt} = \begin{Bmatrix} \left[-B_j^U \Lambda_{jkt} + \left(k_1^U B_j^U + r \right) \right] \Xi_{jkt} - B_j^U \nu_{jkt} \Lambda_{jkt} - \left[W_t - A_j^U + k_2^U B_j^U \right] \Lambda_{jkt} \\ + k_1^U B_j^U \nu_{jkt} - k_1^U A_j^U + k_1^U k_2^U B_j^U \end{Bmatrix} dt,$$

$$\lim_{t \to \infty} \Xi_{jkt} = 0,$$

which given the solution of the above Riccati for $\Lambda_{jk\square}$ is a backward linear first order ODE.

Substituting the linear transformation to the forward equation for $W_{j\square}^U$ we have

$$dW_{jt}^U = \left\{ \left[-k_1^U B_j^U + B_j^U \Lambda_{jkt} \right] W_{jt}^U + B_j^U \Xi_{jkt} + B_j^U \nu_{jkt} + W_t - A_j^U + k_2^U B_j^U \right\} dt,$$

$$W_{00}^U = w_0^U,$$

which is a forward SDE.

Then the backward variable $\mu_{jk\square}$ follows from the linear transformation.

Given now that $W_{j\square}^U$ and $\mu_{jk\square}$ are also known variables we impose again a similar linear transformation

$$\lambda_{jkt}^D = \Pi_{jkt} S_{jkt} + \Sigma_{jkt},$$

where $\Pi_{jk\square}$ and $\Sigma_{jk\square}$ are functions to be determined.

Taking differentials, we have

$$\begin{aligned} d\lambda_{jkt}^D &= \Pi_{jkt} dS_{jkt} + S_{jkt} d\Pi_{jkt} + d\Sigma_{jkt} \\ &= S_{jkt} d\Pi_{jkt} + d\Sigma_{jkt} + \begin{Bmatrix} \left[B_k^D \Pi_{jkt}^2 - k_1^D B_k^D \Pi_{jkt} \right] S_{jkt} + B_k^D \Pi_{jkt} \Sigma_{jkt} - k_1^U B_j^U W_j^U \Pi_{jkt} \\ + \left[B_j^U \mu_{jkt} + B_j^U \nu_{jkt} + W_t - A_j^U + k_2^U B_j^U + R_t - O_t - A_k^D + k_2^D B_k^D \right] \Pi_{jkt} \end{Bmatrix} dt, \end{aligned}$$

while the relationship for $\lambda_{jk\square}^D$ may be written equivalently

$$d\lambda_{jkt}^D = \left\{ \left[-\left(k_1^D \right)^2 B_k^D + \left(r + k_1^D B_k^D \right) \Pi_{jkt} \right] S_{jkt} + \left(r + k_1^D B_k^D \right) \Sigma_{jkt} - k_1^D A_k^D + k_1^D k_2^D B_k^D \right\} dt..$$

A sufficient condition for the latter to be equivalent is

$$d\Pi_{jkt} = \left[-B_k^D \left(\Pi_{jk}^U \right)^2 + \left(2k_1^D B_k^D + r \right) \Pi_{jkt}^U - \left(k_1^D \right)^2 B_k^D \right] dt,$$

$$\lim_{t \to \infty} \Pi_{jkt} = 0,$$

which is a backward Riccati that can be solved numerically.

And

$$
d\Sigma_{jkt} = \left\{ \begin{array}{l} \left[-B_k^D \Pi_{jkt} + \left(k_1^D B_k^D + r \right) \right] \Sigma_{jkt} \\ -\left[-k_1^U B_j^U W_j^U + B_j^U \mu_{jkt} + B_j^U \nu_{jkt} + W_t - A_j^U + k_2^U B_j^U + R_t - O_t - A_k^D + k_2^D B_k^D \right] \Pi_{jkt} \\ -k_1^D A_k^D + k_1^D k_2^D B_k^D \end{array} \right\} dt,
$$

$$
\lim_{t \to \infty} \Sigma_{jkt} = 0,
$$

which given the solution of the above Riccati for $\Pi_{jk\cdot}$ is a backward linear first order ODE.

Substituting the linear transformation to the forward equation for $S_{jk\cdot}$ we have

$$
dS_{jkt} = \left\{ \begin{array}{l} B_k^D \left[\Pi_{jkt} - k_1^D \right] S_{jkt} + B_k^D \Sigma_{jkt} - k_1^U B_j^U W_{jt}^U + B_j^U \mu_{jkt} + B_j^U \nu_{jkt} \\ +W_t - A_j^U + k_2^U B_j^U + R_t - O_t - A_k^D + k_2^D B_k^D \end{array} \right\} dt, \qquad (38)
$$

$$
S_{000} = s_0.
$$

Then the backward variable $\lambda_{jk\cdot}^D$ follows from the linear transformation.

In the following section, we aim to estimate the main components of our model using a stochastic frontier model and a quadratic production function, the form of which remains unknown. In order to identify the production function, we need to estimate the sample coefficients of its main variables, which in order to be a good reflection of their real values, they need to be unbiased, consistent and efficient.

2. THE ECONOMETRIC MODEL

Estimation of production functions has always been a difficult exercise. The reason is that inputs like capital and labor are correlated with the error term for at least three reasons. First, decisions about *inputs* depend on overall productivity. The second source is measurement errors in the right-hand-side variables. The third source is from profit maximization, i.e., the firms choose inputs and output simultaneously to maximize profit.

Olley and Pakes [OP] (1996) use investment as a proxy for such unobservable shocks, while Levinsohn and Petrin [LP] (2003) use intermediate inputs as a better proxy that may respond more smoothly to unobserved productivity shocks. Both approaches which are widely used in the literature have been subject to criticism. This paper presents a new estimation method of firm-level productivity dealing with the endogeneity problem which is pervasive in production function estimation.

Neither OP nor LP are devoid of problems (see Gandhi, Navarro, and Rivers[GNR] (2017), Ackerberg, Caves and Fraser [ACF] (2015), and Doralzeski and Jaumendreu (2013)). GNR (2017) show that, besides the collinearity problem pointed out by ACF (2015), both the OP and LP estimators suffer from the lack of relevant instruments for the endogenous inputs in the model. Ackerberg, Caves and Frazer (2006) have shown that the OP and LP approaches to estimating TFP have a problem of collinearity if labor and intermediate inputs depend on TFP just like investment. GNR (2017) also propose a nonparametric treatment of the production function. There could also be non-linearity due to capital-labour substitution in the sense that when labour input is costly, capital could be substituted to replace labour, making the relationship endogenous and non-linear.

In a substantive extension of the model, we introduce technical inefficiency in production and we allow for autocorrelated TFP. Bayesian analysis is performed using a Sequential Monte Carlo / Particle-Filtering approach.

Consider the following stochastic frontier model:

$$y_{it} = j\,(x_{it}, z_{it}; b) + v_{it} - u_{it}, \quad i = 1, \mathrm{K}, n, t = 1, \ldots, T \tag{39}$$

where y_{it} is the output of firm i and date t, $j\,()$ is an unknown functional form, z_{it} is a $p \times 1$ vector of exogenous inputs, x_{it} is a $p \times 1$ vector of endogenous inputs, b is a $d \times 1$ vectors of unknown parameters, v_{it} is a symmetric random error, u_{it} is the one-sided random disturbance representing technical inefficiency. We assume that z_{it} is uncorrelated with v_{it} and u_{it} but x_{it} is allowed to be correlated with v_{it} and possibly with u_{it}. This, of course, generates an endogeneity problem. We also assume that u_{it} and v_{it} are independent and leave the form of u_{it} unrestricted. The model can be easily extended to the case of exogenous (environmental) variables are included in the distribution of technical inefficiency (see for example, Battese and Coelli (1995) and Caudill, Ford and Gropper (1995)).

To address the endogeneity problem, we propose an approach which does not require the use of instrumental variables, which can often be weak or unreliable, is based on copula functions to determine the joint distribution of the endogenous regressors and the composed errors that effectively capture the dependency between them.

We first assume that $v_{it} \sim i.i.d.N(0, \sigma_v^2)$ and $u_{it} \sim i.i.d.\left|N(0, \sigma_u^2)\right|$. Then the density of $\varepsilon_{it} = v_{it} - u_{it} = y_{it} - \varphi(x_{it}, z_{it}; \beta)$ is given by:

$$g(\varepsilon_{it}) = \int_0^\infty f_v(\varepsilon_{it} + u_{it})\, f_u(u_{it})\, du_{it} = \frac{2}{\sigma}\varphi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(-\frac{\lambda\varepsilon_{it}}{\sigma}\right) \tag{40}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u / \sigma_v$, $\varphi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and cumulative distribution function of a standard normal random variable, respectively. To avoid the non-negativity restrictions we make use of the following transformation: $\bar{\lambda} = \log(\lambda)$ and $\bar{\sigma}^2 = \log(\sigma^2)$. Let $\theta = (\beta', \bar{\lambda}, \bar{\sigma}^2)'$ then it follows that the conditional pdf of y given x and z is

$$f(y_{it} \mid x_{it}, z_{it}) = \frac{2}{\bar{\sigma}}\varphi\left(\frac{y_{it} - \varphi(x_{it}, z_{it}; \beta)}{\bar{\sigma}}\right)\Phi\left(-\frac{\bar{\lambda}}{\bar{\sigma}_v}(y - \varphi(x_{it}, z_{it}; \beta))\right) \tag{41}$$

and conditional log-likelihood is then given by

$$\log L(\theta) = \sum_{i=1}^{n}\sum_{t=0}^{T} \log f(y_{it};\theta \mid x, z). \tag{42}$$

2.1 Copula Approach

In this subsection, we propose an approach that models the dependence between the endogenous regressors and the composed error terms directly via a copula function which does not require the use of instruments. At this stage, e do not introduce dynamic latent productivity, which is left for subsection 2.2. For rigorous treatment on copulas, see for example Nelsen (2006). We take the function $j\,()$ as given and provide its construction in subsection 2.2.

To this end, let $F(x_1, \ldots, x_p, e)$ be the joint distribution of $(x_1, \ldots, x_p)$ and e_i. Now since the information contained in the correlation between $(x_1, \ldots, x_p)$ and e_i is also contained in its joint distribution, and if this is known to belong to a class of parametric density, then consistent estimates of the model parameters can be obtained by simply maximizing the log-likelihood function derived from $F(x_1, \ldots, x_p, e)$. Thus, there is no need for resorting to instruments nor to consistently estimate the parameters of the model.

In practice, however, $F(x_1, \ldots, x_p, e)$ is typically unknown. To address this problem, we follow Park and Gupta (2012) and suggest a copula approach to determine this joint density. The copula essentially captures the dependence in the joint distribution of the endogenous regressors and the composed errors. For exposition purpose, suppose we have a joint distribution of $(x_1,\ldots,x_p,\varepsilon)$ with joint density $f(x_1,\ldots,x_p,\varepsilon)$, and let $f_j(x_j)$, $F_j(x_j)$, for $j = 1,\ldots, p$, $g(\varepsilon)$ and $G(\varepsilon)$ denote the marginal density and CDF of x_j and ε, respectively.

Also, let C denotes the "copula function" defined for $(\xi_1,\ldots,\xi_{p+1}) \in [0,1]^{p+1}$ by
$C(\xi_1,\ldots,\xi_{p+1}) = P(F_1(x_1) \le \xi_1,\ldots,F_p(x_p) \le \xi_p, G(\varepsilon) \le \xi_{p+1})$

, so that the copula function is itself a CDF.

Moreover, since $F_j(x_j)$ and $G(\cdot)$ are marginal distribution function, each component $U_j = F_j(x_j)$ and $U_\varepsilon = G(\varepsilon)$ has a uniform marginal distribution (see for example *Li and Racine*, 2007 in Theorem A.2). Let $c(\xi_1,\ldots,\xi_p)$ denotes the pdf associated with $C(\xi_1,\ldots,\xi_p)$, then by Sklar's theorem (*Sklar*,1959), we have

$$f(x_1,\ldots,x_p,\varepsilon) = c\left(F_1(x_1),\ldots,F_p(x_p),G(\varepsilon)\right) g(\varepsilon) \prod_{j=1}^{p} f_j(x_j). \tag{43}$$

Thus, equation (41) shows that the copula function completely characterizes the dependence structure of $(x_1, \ldots, x_p, e)$, and $c(x_1, \ldots, x_p) = 1$ if and only if $(x_1, \ldots, x_p, e)$ are independent of each other.

To obtain the joint density in (41), we need to specify the copula function. One commonly used copula function is the Gaussian copula. Other copula functions such as Frank, Placket, Clayton, and Farlie-Gumbel-Morgenstern can also be used. The Gaussian copula is generally robust for most application (*Song*, 2000) and has many desirable properties (*Danaher and Smith*, 2011).

Let $\Phi_{\Sigma,p+1}$ denote a $(p+1)$-dimensional CDF with zero mean and correlation matrix Σ. Then the $(p+1)$-dimensional CDF with correlation matrix Σ is given by

$$C(w;\Sigma)=\Phi_{\Sigma,p+1}\left(\Phi^{-1}(U_1),\ldots,\Phi^{-1}(U_p),\Phi^{-1}(U_\varepsilon)\right)$$

, where $w = (U_1,\mathrm{K},U_p,U_e) = (F_1(x_1),\mathrm{K},F_p(x_p),G(e))$.

The copula density is:

$$c(w;\Sigma)=\left(\det(\Sigma)\right)^{-1/2}\times \exp\left\{-\frac{1}{2}\left(\Phi^{-1}(U_1),\ldots,\Phi^{-1}(U_p),\Phi^{-1}(U_\varepsilon)\right)'\left(\Sigma^{-1}-I_{p+1}\right)\left(\Phi^{-1}(U_1),\ldots,\Phi^{-1}(U_p),\Phi^{-1}(U_\varepsilon)\right)\right\}. \quad (44)$$

The log-likelihood function corresponding to (3) is:

$$\log L(\theta,\Sigma)=\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\ln c(F_1(x_{1,it}),\ldots,F_p(x_{p,it}),G(\varepsilon_{it};\theta);\Sigma)+\sum_{j=1}^{p}\ln f_j(x_{j,it})+\ln g(\varepsilon_{it};\theta)\right\}, \quad (45)$$

where $\theta=(\beta',\bar{\lambda},\bar{\sigma}^2)'$ and the form of $c(.)$ is given in (42). Notice that the first term in the summation in (45) is derived from the copula density and this term reflects the dependence between the endogenous variables and the composed errors. In addition, since the marginal density $f_j(x_j)$ does not contain any parameters of interest, the second term in the summation in (45) can be dropped from the log-likelihood function. Finally, it is clear from (43) that if there is no endogeneity problem, (45) collapses to the log-likelihood function of the standard stochastic frontier models.

By maximizing the log-likelihood function in (43), consistent estimates of (θ,Σ) can be obtained, and this can be done as we describe below.

1. Estimation of $F_j(x_j),\ j = 1,\mathrm{K},p$; and $G(\varepsilon;\theta)$

Since $F_j(x_{ji})$ are unknown and we have an observed sample of $x_{ji},\ j = 1,\mathrm{K},p;\ i = 1,\mathrm{K},n$; in the first step, we can estimate $F_j(x_{ji})$ by

$$\tilde{F}_{nj}=\frac{1}{nT+1}\sum_{i=1}^{n}1(x_{j,it}\le x_{0j}),\quad j=1,\ldots,p, \quad (46)$$

where $1(.)$ is an indicator function. Note that in (8), we have used the rescaling factor $1/(nT+1)$ rather than $1/nT$ to avoid difficulties arising from the potential unboundedness of the $\ln c\left(F_1(x_{1,it}),\ldots,F_p(x_{p,it}),G(\varepsilon_{it};\theta);\Sigma\right)$ as some of the $F_j(x_j)$ tend to one. To estimate $G(e_{it};q)$, note that its density $g(\varepsilon_{it};\theta)$ is given in (1) and by definition, $G(\varepsilon_{it};\theta)=\int_{-\infty}^{\varepsilon_{it}} g(s;\theta)ds$, thus $G(\varepsilon;\theta)$ can be estimated using numerical integration, and let denotes the estimator of $G(\varepsilon;\theta)$.

2. Maximization of the log-likelihood function

Maximization of the log-likelihood function in (5) with $F_j(x_j)$ and $G(\varepsilon_{it};\theta)$ are replaced by their $\tilde{G}(\varepsilon_i;\theta)$ estimates $\tilde{G}(\varepsilon_i;\theta)$ $\tilde{F}_j(x_j)$ and , respectively, i.e.,

$$(\hat{\theta},\hat{\Sigma})=\arg\max_{\theta\in\Theta,\Sigma}\sum_{i=1}^{n}\left\{\ln c(\tilde{F}_1(x_{1i}),\ldots,\tilde{F}_p(x_{pi}),\tilde{G}(\varepsilon_i;\theta);\Sigma)+\ln g(\varepsilon_i;\theta)\right\} \quad (47)$$

3. Estimating Technical Inefficiency

Once the parameters have been estimated, the ultimate goal is to predict the values of the technical inefficiency term u_i, and this can be calculated based on *Jondrow et al.* (1982):

$$\hat{u}_{it}=\hat{E}(u_{it}\mid\varepsilon_{it})=\frac{\hat{\sigma}\hat{\lambda}}{1+\hat{\lambda}^2}\left[\frac{\varphi(\hat{\lambda}\hat{\varepsilon}_{it}/\hat{\sigma})}{1-\Phi(\hat{\lambda}\hat{\varepsilon}_{it}/\hat{\sigma})}-\frac{\hat{\lambda}\hat{\varepsilon}_{it}}{\hat{\sigma}}\right], \quad (48)$$

where $\hat{\varepsilon}_{it}=y_{it}-\varphi(x_{it},z_{it};\hat{\beta})$ and $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\sigma}^2$ are the parameter estimates obtained from the approach discussed above.

2.2 LOCAL LIKELIHOOD ESTIMATION

The functional form $\varphi(x_{it},z_{it};\beta)$ was left unspecified so far. Of course, any parametric form can be used but here we focus on non-parametric estimation by the local likelihood method. We use the simpler notation $\varphi(x_{it};\beta)$ as the extension to the case of exogenous covariates is straightforward. Since we have a multivariate covariate, we use the method of local linear estimation. This means that all parameters of the model become functions of x, and they are denoted by $\theta(x)$. We denote the conditional density of y given x by $p(y\mid x)=g(y;\theta(x))$, where $\theta(x)\in\mathbb{R}^k$ is unknown and we define $q(y;\theta(x))=\log g(y;\theta(x))$. For example, a standard frontier would take the form:

$$y_{it}=m(x_{it})+v_{it}-u_{it}, \quad (49)$$

where $v_{it}\mid x_{it}\sim N\left(0,\sigma_v^2(x_{it})\right)$, $u_{it}\mid x_{it}\sim N\left(0,\sigma_u^2(x_{it})\right)$. Then we have

$$\theta(x)=\left[m(x),\sigma_v^2(x),\sigma_u^2(x)\right]'.$$

Our fundamental departure from the standard model is the **introduction of productivity**:

$$y_{it} = m(x_{it}, z_{it}) + v_{it} + \omega_{it} - u_{it}, \tag{50}$$

where the productivity process is as follows:

$$\omega_{it} \mid x_{it}, \omega_{i,t-1} \sim N\left(r(\omega_{i,t-1}, x_{it}, z_{it}), \sigma_{\omega}^2(\omega_{i,t-1}, x_{it}, z_{it})\right).$$

In this specification, $r(\omega_{i,t-1}, x_{it}, z_{it})$ is a non-parametric productivity mean process, and $\sigma_{\omega}^2(\omega_{i,t-1}, x_{it}, z_{it})$ is the variance. For ease in notation we omit explicit dependence on z and we continue to denote $\theta\left(x\right) \in \mathbb{R}^k$ with

$$\theta(x) = \left[m(x), r(\omega_{-1}, x), \sigma_v^2(x), \sigma_u^2(x), \sigma_\omega^2(\omega_{-1}, x)\right]',$$

where w_{-1} denotes the lagged value of productivity. As productivity is latent special problems are introduced into the analysis.
There is a multivariate kernel which satisfies:

$$\int K(u)du = 1, \quad \int uu'K(u)du = \mu_2 I_d.$$

To fix notation, we start with the analysis of the simpler model in (11). The conditional local linear log-likelihood is given by[9]

$$\log L(\theta_o, \Theta_1) = \sum_{i=1}^{n} \sum_{t=1}^{T} q(y_{it}, \theta_o + \Theta_1\left(x_{it} - x\right)) K_H\left(x_{it} - x\right), \tag{51}$$

where θ_o, Θ_1 is a vector ($k \times 1$) and matrix ($k \times d$) respectively, H is a bandwidth matrix which is symmetric, positive definite and $K_H\left(u\right) = \left|H\right|^{-1} K\left(H^{-1}u\right)$. We choose a multivariate product kernel so that $K(u) = \prod_{j=1}^{d} K_j(u_j)$ in which case

$$\int uu'K(u)du = \left(\int u_1^2 K_1(u_1)du_1\right) I_d.$$

The local linear estimator is $\hat{\theta}(x) = \hat{\theta}_o(x)$ where $\hat{\theta}_o(x)$ and $\hat{\Theta}_1\left(x\right)$ maximize the log-likelihood $L(\theta_o, \Theta_1)$ with respect to θ_o, Θ_1. Computational details are in *Kumbhakar et al.* (2007), (Sections 3.1 and 3.2) and we follow this paper closely.

For the model with latent productivity w_{it} as in (12) the likelihood function is

$$L(\theta_o, \Theta_1) = \int_{\mathbb{R}^{nT}} \left\{\prod_{i=1}^{n} \prod_{t=1}^{T} g(y_{it}, \omega_{it}, \theta_o + \Theta_1\left(\Lambda_{it} - \Lambda\right)) \cdot K_H\left(\Lambda_{it} - \Lambda\right)\right\} d\omega, \tag{52}$$

where $\Lambda_{it} = \left[x_{it}', \omega_{i,t-1}\right]'$, $\Lambda = [x', \omega_{-1}]'$, and

[9] In fact we include z_{it} in the kernel functions because, in this instance, they represent important environmental variables that help in modeling heterogeneity. For ease in notation we redefine x=[x', z']'.

$$g\left(y,\omega;\theta(\Lambda)\right)=\frac{2}{\sigma(x)}\varphi\left(\frac{y_{it}-\varphi(x_{it};\beta(x))-\omega_{it}}{\sigma(x)}\right)\Phi\left(-\frac{\lambda(x)}{\sigma_v(x)}(y-\varphi(x_{it};\beta(x)))-\omega_{it}\right)$$
$$\cdot\frac{1}{\sigma_\omega(x,\omega_{-1})}\varphi\left(\frac{\omega_{it}-r(x_{it},\omega_{i,t-1};\gamma(x,\omega_{-1}))}{\sigma_\omega(x,\omega_{-1})}\right). \quad (53)$$

Moreover, $\gamma\left(x,\omega_{-1}\right)$ denotes the localized parameters in the r() function in (12). For ease in notation we define $\theta\left(x,\omega_{-1}\right)=\left[\beta\left(x\right)',\gamma\left(x,\omega_{-1}\right)'\right]' \in \mathbb{R}^k$.

In (52) there is an nT - dimensional integral which we cannot evaluate analytically, which is obvious from the definition of (53). The computation relies in two steps.

Step 1: Integrate out $\left\{\omega_{it}\right\}$ from (14) using a Sequential Monte Carlo algorithm (Pitt and Shephard, 1997).

Step 2: Maximize the resulting expression using numerical optimization techniques.

For reasons of computational convenience and without sacrificing generality we assume

$\omega_{it}=\rho\omega_{i,t-1}+\xi_{it},\ \left\{\xi_{it}\right\}\sim i.i.d.\ N(0,\sigma_\xi^2).$ (54)We will still need the SMC algorithm in step 1. For the SMC algorithm we use 10^6 particles per likelihood evaluation and a standard conjugate gradients algorithm for maximization. Our results were insensitive to using 105 or 107 particles per likelihood evaluation.

3. EMPIRICAL RESULTS

In this section we perform a simple nonparametric estimation of the production function per sector in each country, Ethiopia and Kenya, based on the econometric model described in Subsection 3.3.1. Our dataset includes 16 observations for each sector of these countries from 2000 to 2015. This methodology takes into account the regional differences in productivity between the upstream and downstream countries, which is a necessary categorization of these countries due to the formulation of the water resources management problem investigated in Section 1.2.

Data preparation is a critical first step for building high performance predictive models. At first, we convert all the monetary variables to constant 2010 prices, since the prices we had available for each year of our period could not be used for comparisons thanks to inflation effects. Additionally, we perform all the necessary transformations of the variables to end up either with the same units of measurement or with suitably scaled data, so as to standardize predictions subject to the units of the regression coefficients.

The results of the nonparametric estimation are reported in Table 1 and Table 2. Input and output variables were transformed to their corresponding log values and were normalized by their respective sample means. From the estimated production function for each of the two countries we can easily obtain their corresponding marginal product function, which is connected with the water use input variable via the relationship:

$$\text{Marginal product} = \alpha + \beta \cdot \text{water use} \quad (55)$$

Both coefficients for each country turn out to have the expected signs. As explained at section 2.2, inverse demand function and it is expected to be equal to the marginal contribution of the water to the output of each sector given by equation (55). Consequently, the derived demand curve for water of the producer is represented at equation (56) showing producer's demand for an input, i.e. the water, as a result of the demand for another related good, i.e. energy.

$$\text{Water use} = a + b \cdot \text{price} \quad (56)$$

where a is the intercept of water demand of each sector and b the price elasticity of water demand.

In order to calculate the price elasticities, we used the formula (57). In alignment with Figure 1, it is noticed that all sectors are exceptionally inelastic to a price change for water use, i.e. relatively large changes in price cause very small changes in demanded quantity of water. In particular, Agriculture seems to be perfectly inelastic to any price change, which means that in both countries the demanded quantity will remain stable for any price change and so the price cannot influence the water use. This implies an extremely strong relationship between the input, in this case water, and the corresponding output of each sector such as the seed-producing crops, since the producer lacks alternatives and so values highly the use of water. The elasticities for both countries are presented in Table 3.

$$\text{price elasticity} = \frac{d(\text{water use})}{d(\text{price})} \cdot \frac{\text{price}}{\text{water use}} \quad (57)$$

Table 1 - Empirical results: β parameter for each Sector

	Mining	**Energy**	**Tourism**	**Residential**	**Agriculture**
Ethiopia	-0.0010	-0.0014	-0.0012	-0.0013	-0.0000321
Kenya	-0.0011	-0.0013	-0.0010	-0.0015	-0.0000319

Table 2 - Empirical results: α parameter for each Sector

	Mining	**Energy**	**Tourism**	**Residential**	**Agriculture**
Ethiopia	1.80	1.73	1.48	1.65	1.48
Kenya	1.54	1.70	1.56	1.77	1.56

Table 3 - Price elasticity for each Sector

	Mining	**Energy**	**Tourism**	**Residential**	**Agriculture**
Ethiopia	-0.099	-0.131	-0.096	-0.116	-0.003
Kenya	-0.092	-0.120	-0.085	-0.143	-0.003

Finally, the different demand curves, coming from equation (56) for all 5 sectors of the OTB riparian countries, provide us with an ordering of these sectors via their demand function intercepts. Sequential exits from the market are defined by the relative importance of sector-specific demand parameter ratio a , with $a = \alpha/\beta$. As water demand for each of these economic sectors reaches zero sequentially, its price increases revealing so producers' preferences for water use. At these prices, in Ethiopia Tourism sector should exit the market first followed by the Residential and the Energy sector, while in Kenya Mining sector would exit the market first trailed by the Tourism sector. However, in both cases, in case of river/lake depletion, agriculture sector should be the last one to exit the market, since it is valuing water use more than any other sector.

Figure 1 illustrates the derived demand for water of each sector in each country. In both cases, the agriculture sector is almost inelastic in water use declaring so, an intense connection between water use and crops, which is caused by the existence of large irrigation schemes in the basin. Moreover, producers in mining sector in Kenya values higher the water than in Ethiopia, and that happens because Kenya relies strongly on groundwater for mining production.

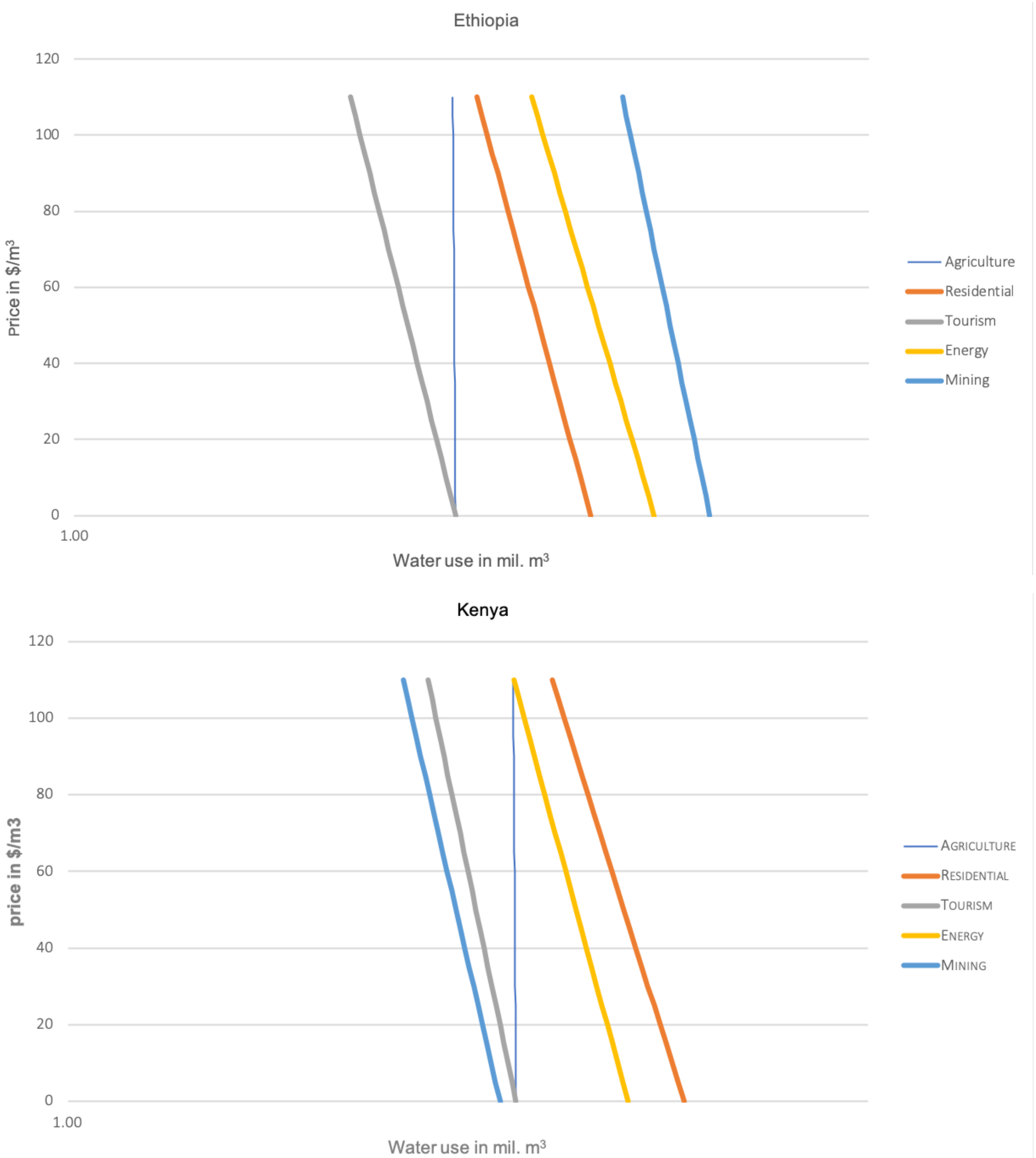

Figure 1 - Curves of the demand function of All Sectors for the Upstream and Downstream Country

As presented at Figure 2, sampling distributions of water elasticities tend to not vary significantly between them. In particular, only the distribution for the residential sector in Ethiopia (upper panel) is shown to look like Normal distribution, while the others show disorders at their extreme cases. None of these means is the mode of the distribution as well, although the chasm between those values is not notable. In economic terms, the elasticities for water demand in each sector do not deviate remarkably, letting so similar behavioural patterns to be observed in each sector across the two countries of interest.

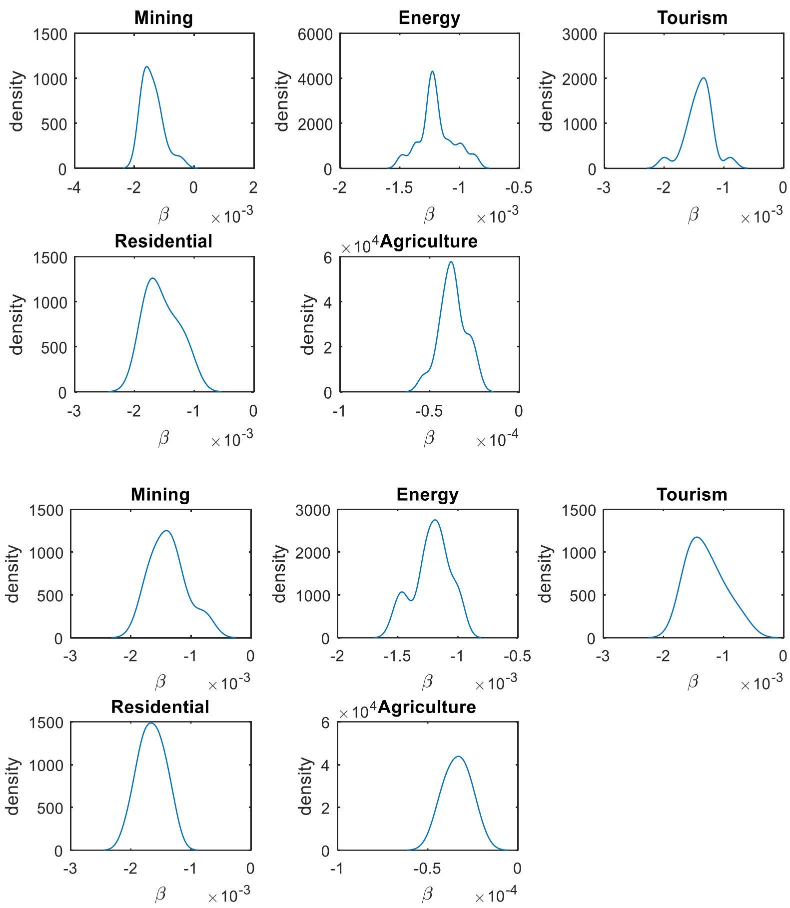

Figure 2 - Sampling distributions of water elasticities by sector for Ethiopia (upper panel) and Kenya (lower panel)

The second parameter of the inverse demand curve is the constant term, which is responsible for the starting point of the demand curve, revealing so the willingness to pay (WTP) of the stakeholders in each sector. Figure 3 shows the distributions of constant terms of the inverse demand functions and interestingly we can see that in most cases the WTP for water use in energy sector is greater than the corresponding one in agriculture and tourism, which implies greater profitability in energy sector. Additionally, in terms of WTP, mining sector in Ethiopia, which follows a leptokurtic distribution seems to be the most stable one.

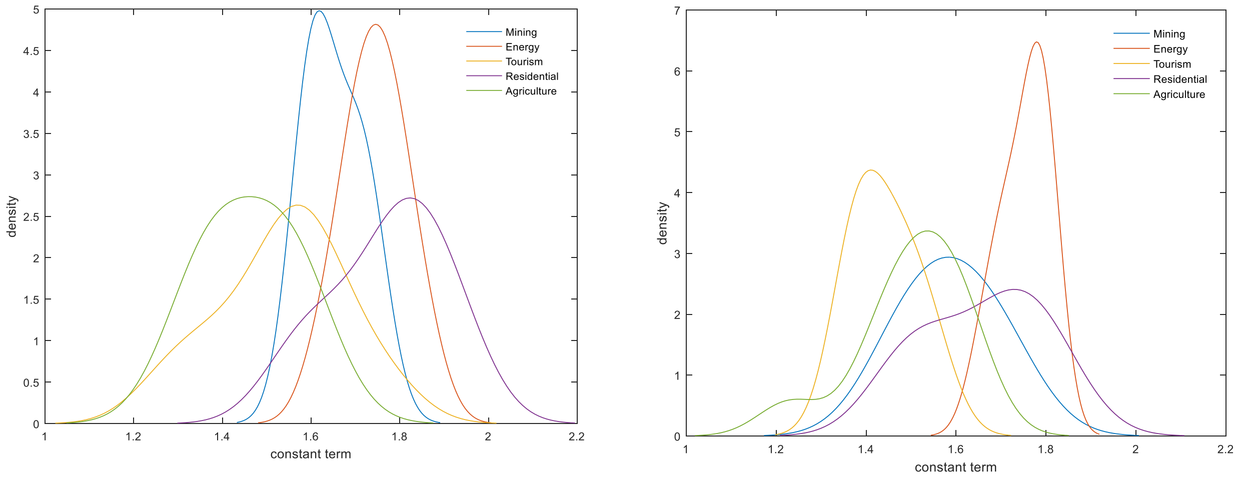

Figure 3 - Sampling distributions of constant terms by sector for Ethiopia (left) and Kenya (right)

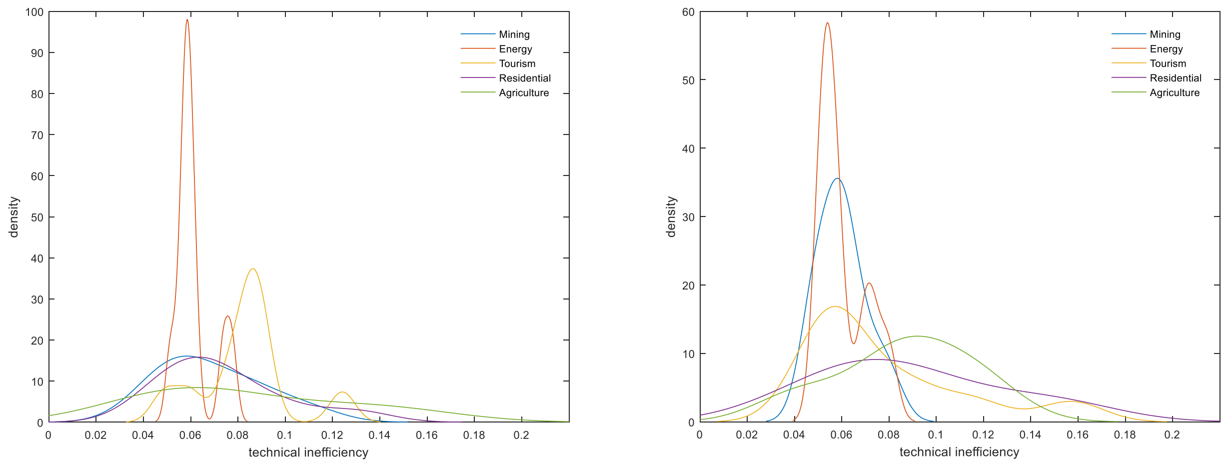

Figure 4- Sampling distributions of technical inefficiency by sector for Ethiopia (left) and Kenya (right)

Figure 4 presents technical inefficiency parameter by sector for the two countries of interest. A zestful outcome is the fact that Mining and Residential sectors in Ethiopia follow exactly the same distribution with a positive skew to the right. In Energy and Tourism sector in both countries, u_{it} has two district peaks (bimodal distribution), which indicates that in these sectors there are two groups of producers, some of them achieve to maximize their outputs given their inputs, while some others do not with technical inefficiency taking greater values than the former group. However, it is noteworthy that Energy sector is more technical efficient in comparison with Tourism, since the lowest peak of Tourism is as great at the biggest one of Energy sector

Another important indicator is the technical change by sector as presented in Figure 5. Residential, Agriculture, Energy and Tourism in Ethiopia. distributions resemble Normal distribution, while Mining sector’s distribution has two peaks. However, in Kenya the peak of most distributions is in zero declaring so, that the majority of sectors remains stable without being engaged to innovative changes.

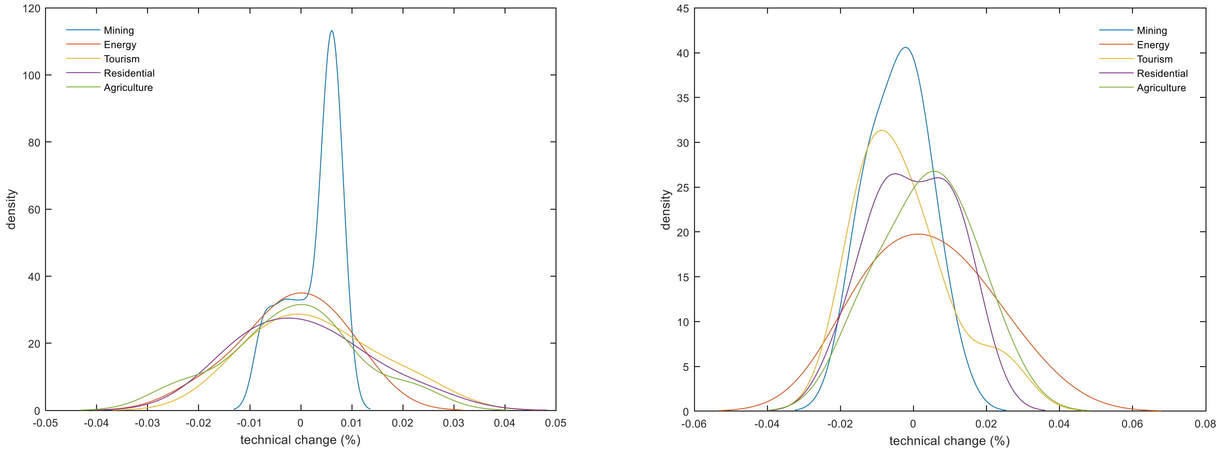

Figure 5 - Sampling distributions of technical change by sector for Ethiopia (left) and Kenya (right)

Figure 6 presents productivity growth by sector with more particular case the multimodal distribution of the Residential sector in Ethiopia. In this case, there are three district peaks, with zero growth rate. Agriculture sector in Ethiopia also formulates two peaks, with the most common having zero mean as well revealing so lack in developing new technologies and making so production more efficient. Additionally, the Tourism sector of Kenya also formulates two peaks, with the most common one lying in the positive side, which underlines the development advantage of the Tourism sector in comparison with the other sectors in both countries.

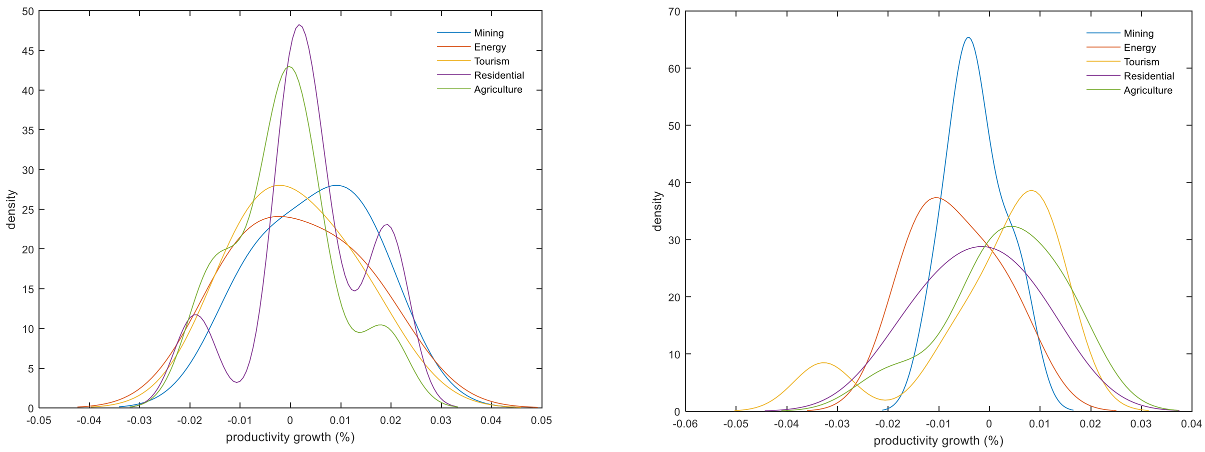

Figure 6 - Sampling distributions of productivity growth (%) by sector for Ethiopia (left) and Kenya (right)

REFERENCES

Ackerberg, D. A., Caves, K. and G. Frazer (2015) Identification Properties of Recent Production Function Estimators. *Econometrica*, 83, 2411-2451.

Aigner, D.J., Lovell, C.A.K. and P. Schmidt (1977) Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, 6 (1), 21-27.

Amsler, C., A. Prokhorov and P. Schmidt (2014a) Using copula to model time dependence in stochastic frontier models. *Econometric Reviews*, 33(5-6), 497-522.

Amsler, C., A. Prokhorov and P. Schmidt (2014b) Endogeneity in stochastic frontier models. *Working Paper*, Michigan State University.

Andrieu, C., and G.O. Roberts (2009) The pseudo-marginal approach for efficient computation. *Ann. Statist.*, 37, 697–725.

Andrieu, C., A. Doucet, and R. Holenstein (2010) Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72, 269–342.

Bhaduri, A., Manna, U. and E. Barbier (2011) Climate change and cooperation in transboundary water sharing: an application of stochastic Stackelberg differential game s in Volta river basin*, Natural Resource Modeling*, 24(4), 409-444.

Battese. G.E. and T.J. Coelli (1995) A model for technical inefficiency effects in a stoschastic frontier production function for panel data. *Empirical Economics*, 20, 325-332.

Cappı, O., E. Moulines, E., and T. Rydın (2005) Inference in Hidden Markov Models. Springer, Berlin.

Casarin, R., J.-M. Marin (2007) Online data processing: Comparison of Bayesian regularized particle filters. University of Brescia, Department of Economics. *Working Paper n. 0704.*

Caudill, S.B., J.M. Ford and D.M. Gropper (1993) Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity. *Journal of Business Economic Statistics*, 13, 105-111.

Dagenais, M.G. and D.L. Degenais (1997) Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics* 76 (1-2), 193-221.

Danaher, P.J. and M.S. Smith (2011) Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30 (1), 4-21.

Doraszelski, U., and J. Jaumandreu (2013) R&D and Productivity: Estimating Endogenous Productivity, *Review of Economic Studies*, 80, 1338 - 1383.

Doucet, A., N. de Freitas, and N. Gordon (2001) Sequential Monte Carlo Methods in Practice. New York: Springer.

Fearnhead, P. (2007). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statistics and Computing*, 18(2), 151-171.

Flury, T., and N. Shephard (2011) Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models*. Econometric Theory*, 27, 933-956.

Gordon, N.J. (1997). A hybrid bootstrap filter for target tracking in clutter*. IEEE Transactions on Aerospace and Electronic Systems*, 33, 353– 358.

Gordon, N.J., D.J. Salmond and A.F.M. Smith (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE-Proceedings-F 140: 107–113.

Hall, J., M.K. Pitt, and R. Kohn (2014). Bayesian inference for nonlinear structural time series models. *Journal of Econometrics*, 179 (2), 99–111.

Hausman, J.A., W.K. Newey, and J.L. Powell (1995), Nonlinear errors in variables: Estimation of some Engel curves. *Journal of Econometrics*, 65, 205-233.

Heckman, J.J. and B.E. Honore (1989) The identifiability of the competing risks model. *Biometrika*, 76 (2), 325-330.

Jondrow, J., C.A.K. Lovell, I.S. Materov and P. Schmidt (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19 (2/3), 233-238.

Kim, C. S., Moore, M. R. and J. J. Hanchar (1989) A Dynamic Model of Adaptation to Resource Depletion: Theory and an Application to Groundwater Mining, *Journal of Environmental Economics and Management*, 17, 66-82.

Koetter, M., and O. Vins (2008) The quiet life hypothesis in banking: evidence from German savings banks. *Working paper*.

Koetter, M., J. Kolari, and L. Spierdijk (2012) Enjoying the Quiet Life under Deregulation?: Evidence from Adjusted Lerner Indices for U.S. Banks. *Review of Economics and Statistics,* 94(2), 462-480.

Koundouri, P. and C. Christou (2006) Dynamic Adaptation to Resource Scarcity and Backstop Availability: Theory and Application to Groundwater. *The Australian Journal of Agricultural and Resources Economics*, 50, 227-245.

Kumbhakar, S., Parmeter, C.F. and E.G. Tsionas (2013) A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172, 66-76.

Kutlu, L. (2010) Battese-Coelli estimator with endogenous regressors. *Economics Letters*, 109, 79-81.

Lee, L.F. (1983) Generalized econometric models with selectivity. *Econometrica*, 61, 381-428.

Lewbel, A. (1997) Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica*, 65 (5), 1201-1213.

Liu, J. and M. West (2001) Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag.

Levinsohn, J. and A. Petrin (2003) Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70 (2), 317-341.

Li, Q. and J. Racine, (2007) *Nonparametric Econometrics*. Princeton University Press, Princeton, NJ.

Meeusen, W. and J. van den Broeck (1997) Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18 (2), 435-444.

Nelsen, R.B. (2006) *An Introduction to Copula*, Vol. 139 of Springer Series in Statistics, Springer.

Nemeth, C. and P. Fearnhead (2014) Particle Metropolis adjusted Langevin algorithms for state-space models. Pre-print arXiv:1402.0694v1.

Olley, S. and A. Pakes (1996) The dynamics of productivity in the telecommunications equipment industry, *Econometrica*, 64, 1263-1297.

Park, S. and S. Gupta (2012) Handling endogenous regressors by joint estimation using Copulas. *Marketing Science*, 31 (4), 567-586.

Pitt, M.K., N. Shephard (1999). Filtering via simulation based on auxiliary particle filters. *Journal of the American Statistical Association,* 94 (446), 590-599

Ristic, B., Arulampalam, S. and N. Gordon (2004) Beyond Kalman Filters: Particle Filters for Applications. Norwood, MA: Artech House.

Sklar, A. (1959) Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Uinversite de Paris,* 8, 229-231.

Song, PX-K (2000) Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 27 (2), 305-320.

Tran, K.C. and E.G. Tsionas (2013) GMM estimation of stochastic frontier models with endogenous regressors. *Economics Letters*, 118, 233-236.

Tran, K.C. and E.G. Tsionas (2014) Zero-inefficiency stochastic frontier models with varying mixing proportion: Semiparametric approach with application to U.S. banks. *Working Paper, University of Lethbridge. Published: EJOR*.

Trivedi, P.K. and D.M. Zimmer (2006) Using trivariate copulas to model sample selection and treatment effects: Application to family healthcare demand. *Journal of Business and Economic Statistics*, 24(1), 63-76.

APPENDIX. SEQUENTIAL MONTE CARLO

The particle filter methodology can be applied to state space models of the general form:

$$y_T \sim p(y_t \mid x_t),\ s_t \sim p(s_t \mid s_{t-1}), \tag{1}$$

where s_t is a state variable. For general introductions see Gordon (1997), Gordon et al. (1993), Doucet et al (2001) and Ristic et al. (2004).

Given the data Y_t the posterior distribution $p(s_t \mid Y_t)$ can be approximated by a set of (auxiliary) particles $\left\{s_t^{(i)}, i=1,...,.N\right\}$ with probability weights $\left\{w_t^{(i)}, i=1,...,N\right\}$ where $\sum_{i=1}^{N} w_t^{(i)} = 1$. The predictive density can be approximated by:

$$p(s_{t+1} \mid Y_t) = \int p(s_{t+1} \mid s_t) p(s_t \mid Y_t) ds_t \simeq \sum_{i=1}^{N} p(s_{t+1} \mid s_t^{(i)}) w_t^{(i)}, \tag{2}$$

and the final approximation for the filtering density is:

$$p(s_{t+1} \mid Y_t) \propto p(y_{t+1} \mid s_{t+1}) p(s_{t+1} \mid Y_t) \simeq p(y_{t+1} \mid s_{t+1}) \sum_{i=1}^{N} p(s_{t+1} \mid s_t^{(i)}) w_t^{(i)}. \tag{3}$$

The basic mechanism of particle filtering rests on propagating $\left\{s_t^{(i)}, w_t^{(i)}, i=1,...,N\right\}$ to the next step, viz. $\left\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i=1,...,N\right\}$ but this often suffers from the weight degeneracy problem. If parameters $\theta \in \Theta \in \Re^k$ are available, as is often the case, we follow Liu and West (2001) parameter learning takes place via a mixture of multivariate normals:

$$p(\theta \mid Y_t) \simeq \sum_{i=1}^{N} w_t^{(i)} N(\theta \mid a\theta_t^{(i)} + (1-a)\overline{\theta}_t, b^2 V_t), \tag{4}$$

where $\overline{\theta}_t = \sum_{i=1}^{N} w_t^{(i)} \theta_t^{(i)}$, and $V_t = \sum_{i=1}^{N} w_t^{(i)} (\theta_t^{(i)} - \overline{\theta}_t)(\theta_t^{(i)} - \overline{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0,1)$ as $a = (1-b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also Casarin and Marin (2007).

Andrieu and Roberts (2009), Flury and Shephard (2011) and Pitt et al. (2012) provide the Particle Metropolis-Hastimgs (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y \mid \theta)$ as $p(Y \mid \theta)$ is often not available in closed form.

Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say $L^j = \hat{p}_N(Y \mid \theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c \mid \theta^{(j)})$ yielding $L^c = \hat{p}_N(Y \mid \theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min\left\{1, \frac{p(\theta^c)L^c}{p(\theta^{(j)})L^j}\frac{q(\theta^{(j)} \mid \theta^c)}{q(\theta^c \mid \theta^{(j)})}\right\}, \tag{5}$$

otherwise we repeat the current draws: $\left\{\theta^{(j+1)}, L^{j+1}\right\} = \left\{\theta^{(j)}, L^j\right\}$.

Hall, Pitt and Kohn (2014) propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering (Pitt et al., 2012) used within PMCMC requires far fewer particles that the standard particle filtering algorithm to approximate $p(Y \mid \theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t \mid s_{t-1})$. When this is not possible, Hall, Pitt and Kohn (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have:

$$p(y_t \mid s_{t-1}) = \int p(y_t \mid s_t) p(s_t \mid s_{t-1}) ds_t, \tag{6}$$

$$p(u_t \mid s_{t-1}; y_t) = p(y_t \mid s_{t-1}, u_t) p(u_t \mid s_{t-1}) / p(y_t \mid s_{t-1}). \tag{7}$$

If one can evaluate $p(y_t \mid s_{t-1})$ and simulate from $p(u_t \mid s_{t-1}; y_t)$ the filter would be fully adaptable (Pitt and Shephard, 1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t \mid s_{t-1})$ by matching the first two moments of $p(y_t \mid s_{t-1})$. So in some way we find that the approximating density $p(y_t \mid s_{t-1}) = N\left(E(y_t \mid s_{t-1}), V(y_t \mid s_{t-1})\right)$. In the second stage, we know that $p(u_t \mid y_t, s_{t-1}) \propto p(y_t \mid s_{t-1}, u_t) p(u_t)$. For $p(u_t \mid y_t, s_{t-1})$ we know it is multimodal so suppose it has M modes are $\hat{u}_t^m$, for $m = 1, \ldots, M$. For each mode we can use a Laplace approximation. Let $l(u_t) = log\left[p(y_t \mid s_{t-1}, u_t) p(u_t)\right]$. From the Laplace approximation we obtain:

$$l(u_t) \approx l(\hat{u}_t^m) + \tfrac{1}{2}(u_t - \hat{u}_t^m)' \nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m). \tag{8}$$

Then we can construct a mixture approximation:

$$g(u_t \mid x_t, s_{t-1}) = \sum_{m=1}^{M} \lambda_m (2\pi)^{-d/2} \mid \Sigma_m \mid^{-1/2} \exp\left\{\tfrac{1}{2}(u_t - \hat{u}_t^m)' \Sigma_m^{-1}(u_t - \hat{u}_t^m\right\}, \tag{9}$$

where $\Sigma_m = -\left[\nabla^2 l(\hat{u}_t^m)\right]^{-1}$ and $\lambda_m \propto \exp\left\{l(u_t^m)\right\}$ with $\sum_{m=1}^{M} = 1$. This is done for each particle s_t^i. This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a proposal for s_t directly from the measurement density $p(y_t \mid s_t)$ although Hall, Pitt and Kohn (2014) are quite right in pointing out that the state equation can be very informative.

In the standard particle filter of Gordon et al. (1993) particles are simulated through the state density $p(s_t^i \mid s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t \mid s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

For $t = 0, \ldots, T-1$ given samples $s_t^k \sim p(s_t \mid Y_{1:t})$ with mass π_t^k for $k = 1, \ldots, N$.

1) For $k = 1, \ldots, N$ compute $\omega_{t|t+1}^k = g(y_{t+1} \mid s_t^k)\pi_t^k$, $\quad \pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^N \omega_{t|t+1}^i$.

2) For $k = 1, \ldots, N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t}^i (ds_t)$.

3) For $k = 1, \ldots, N$ draw $u_{t+1}^k \sim g(u_{t+1} \mid \tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; u_{t+1}^k)$.

4) For $k = 1, \ldots, N$ compute

$$\omega_{t+1}^k = \frac{p(y_{t+1} \mid s_{t+1}^k)\, p(u_{t+1}^k)}{g(y_{t+1} \mid s_t^k)\, g(u_{t+1}^k \mid \tilde{s}_t^k, y_{t+1})}, \qquad \pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}. \tag{10}$$

It should be mentioned that the estimate of likelihood from ADPF is:

$$p(Y_{1:T}) = \prod_{t=1}^T \left(\sum_{i=1}^N \omega_{t-1|t}^i \right) \left(N^{-1} \sum_{i=1}^N \omega_t^i \right). \tag{11}$$

PARTICLE METROPOLIS ADJUSTED LANGEVIN FILTERS

Nemeth et al. (2014) provide a particle version of a Metropolis Adjusted Langevin algorithm (MALA). In Sequential Monte Carlo we are interested in approximating $p(s_t \mid Y_{1:t}, \theta)$. Given that:

$$p(s_t \mid Y_{1:t}, \theta) \propto g(y_t \mid x_t, \theta) \int f(s_t \mid s_{t-1}, \theta)\, p(s_{t-1} \mid y_{1:t-1}, \theta)\, ds_{t-1}, \tag{12}$$

where $p(s_{t-1} \mid y_{1:t-1}, \theta)$ is the posterior as of time $t-1$. If at time $t-1$ we have a set set of particles $\{s_{t-1}^i, i = 1, \ldots, N\}$ and weights $\{w_{t-1}^i, i = 1, \ldots. N\}$ which form a discrete approximation for $p(s_{t-1} \mid y_{1:t-1}, \theta)$ then we have the approximation:

$$\hat{p}(s_{t-1} \mid y_{1:t-1}, \theta) \propto \sum_{i=1}^N w_{t-1}^i f(s_t \mid s_{t-1}^i, \theta). \tag{13}$$

See Andrieu et al. (2010) and Cappe at al. (2005) for reviews. From (13) Fernhead (2007) makes the important observation that the joint probability of sampling particle s_{t-1}^i and state s_t is:

$$\omega_t = \frac{w_{t-1}^i g(y_t \mid s_t, \theta) f(s \mid s_{t-1}^i, \theta)}{\xi_t^i q(s_t \mid s_{t-1}^i, y_t, \theta)}, \tag{14}$$

where $q(s_t \mid s_{t-1}^i, y_t, \theta)$ is a density function amenable to simulation and

$$\xi_t^i q(s_t \mid s_{t-1}^i, y_t, \theta) \simeq c g(y_t \mid s_t, \theta) f(s_t \mid s_{t-1}^i, \theta), \tag{15}$$

and C is the normalizing constant in (12).

In the MALA algorithm of Roberts and Rosenthal (1998)[10] we form a proposal:

$$\theta^c = \theta^{(s)} + \lambda z + \tfrac{\lambda^2}{2} \nabla \log p(\theta^{(s)} \mid Y_{1:T}), \tag{16}$$

where $z \sim N(0, I)$ which should result in larger jumps and better mixing properties, plus lower autocorrelations for a certain scale parameter λ. Acceptance probabilities are:

$$a(\theta^c \mid \theta^{(s)}) = \min \left\{ 1, \frac{p(Y_{1:T} \mid \theta^c) q(\theta^{(s)} \mid \theta^c)}{p(Y_{1:T} \mid \theta^{(s)}) q(\theta^c \mid \theta^{(s)})} \right\}. \tag{17}$$

Using particle filtering it is possible to create an approximation of the score vector using Fisher's identity:

$$\nabla \log p(Y_{1:T} \mid \theta) = E\left[\nabla \log p(s_{1:T}, Y_{1:T} \mid \theta) \mid Y_{1:T}, \theta \right], \tag{18}$$

which corresponds to the expectation of:

$$\nabla \log p(s_{1:T}, Y_{1:T} \mid \theta) = \nabla \log p(\mid s_{1:T-1}, Y_{1:T-1} \mid \theta) + \nabla \log g(y_T \mid s_T, \theta) + \nabla \log f(s_T \mid s \mid_{T-1}, \theta),$$

over the path $s_{1:T}$. The particle approximation to the score vector results from replacing $p(s_{1:T} \mid Y_{1:T}, \theta)$ with a particle approximation $\hat{p}(s_{1:T} \mid Y_{1:T}, \theta)$. With particle i at time t-1 we can associate a value $\alpha_{t-1}^i = \nabla \log p(s_{1:t-1}^i, Y_{1:t-1} \mid \theta)$ which can be updated recursively. As we sample κ_i in the APF (the index of particle at time $t-1$ that is propagated to produce the i th particle at time t) we have the update:

$$\alpha_t^i = a_{t-1}^{\kappa_i} + \nabla \log g(y_t \mid s_t^i, \theta) + \nabla \log f(s_t^i \mid s_{t-1}^i, \theta). \tag{19}$$

To avoid problems with increasing variance of the score estimate $\nabla \log p(Y_{1:t} \mid \theta)$ we can use the approximation:

$$\alpha_{t-1}^i \sim N(m_{t-1}^i, V_{t-1}). \tag{20}$$

The mean is obtained by shrinking α_{t-1}^i towards the mean of α_{t-1} as follows:

---

[10] The benefit of MALA over Random-Walk-Metropolis arises when the number of parameters n is large. This happens because the scaling parameter λ is $O(n^{-1/2})$ for Random-Walk-Metropolis but it is $O(n^{-1/6})$ for MALA, see Roberts et al. (1997) and Roberts and Rosenthal (1998)

$$m_{t-1}^{i} = \delta \alpha_{t-1}^{i} + (1-\delta)\sum_{i=1}^{N} w_{t-1}^{i}\alpha_{t-1}^{i}, \tag{21}$$

where $\delta \in (0,1)$ is a shrinkage parameter. Using Rao-Blackwellization one can avoid sampling α_t^i and instead use the following recursion for the means:

$$m_t^i = \delta m_{t-1}^{\kappa_i} + (1-\delta)\sum_{i=1}^{N} w_{t-1}^{i} m_{t-1}^{i} + \nabla \log g(y_t \mid s_t^i, \theta) + \nabla \log f(s_t^i \mid s_{t-1}^{\kappa_i}, \theta), \tag{22}$$

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t} \mid \theta) = \sum_{i=1}^{N} w_t^i m_t^i. \tag{23}$$

As a rule of thumb Nemeth, Sherlock and Fearnhead (2014) suggest taking $\delta = 0.95$. Furthermore, they show the important result that the algorithm should be tuned to the asymptotically optimal acceptance rate of 15.47% and the number of particles must be selected so that the variance of the estimated log-posterior is about 3. Additionally, if measures are not taken to control the error in the variance of the score vector, there is no gain over a simple random walk proposal.

Of course, the marginal likelihood is:

$$p(Y_{1:T} \mid \theta) = p(y_1 \mid \theta)\prod_{t=2}^{T} p(y_t \mid Y_{1:t-1}, \theta), \tag{24}$$

where

$$p(y_t \mid Y_{1:t-1}, \theta) = \int g(y_t \mid s_t) \int f(s_t \mid s_{t-1}, \theta) p(s_{t-1} \mid Y_{1:T-1}, \theta) ds_{t-1} ds_t, \tag{25}$$

provides, in explicit form, the predictive likelihood.