Dams, Snails and Poverty Traps

Daniele Rinaldo*

Faculty of Economics, University of Cambridge, UK

Abstract

Irrigation schemes are one of the most important policy responses designed to reduce poverty, particularly in sub-Saharan Africa. Concomitantly, they facilitate the propagation of schistosomiasis, a water-based debilitating disease that is endemic in many developing countries. I study the economic impact of schistosomiasis in Burkina Faso via the estimation of its burden on agricultural production with new data and new methods, and identify it as a productivity shock. I show that the disease is both a driver and a consequence of poverty, and that returns to water resources development are significantly reduced once its health effects are taken into account. I reconcile these results with a theoretical framework, which shows how the joint dynamics of schistosomiasis and the production decisions of farmers create Pareto-inferior endemic Nash equilibria, and how the wealth-dependent disease reproduction rate (the R_0) can generate poverty traps. A stochastic extension of the model shows how this rate controls the probability flow between the system attractors. Social optima require deviations from separable allocations proportional to the disease burden on the maximized utility paths. Complete information on the feedback between wealth and disease can potentially allow farmers to escape the poverty trap.
JEL Codes: C61, C73, D15, I15, O13

1 Introduction

Schistosomiasis (*Bilharzia*, or snail fever) is a water-based debilitating neglected tropical disease that affects an estimated 250 million people, more than 85% of whom live in sub-Saharan Africa (Walz et al., 2015). Claiming 3 million disability-adjusted life years per year in the past decade, its global burden ranks second only to malaria among parasitic diseases (James et al., 2018). Severe morbidity due to schistosomiasis results from the accumulation of eggs laid by flatworms of genus *Schistosoma* in the tissues of the human host, leading to a chronic inflammatory response. The parasite species that cause the two main forms of the disease (intestinal and uro-genital) present a complex life cycle involving aquatic snails as intermediate hosts. When untreated, advanced forms of schistosomiasis lead to kidney failure, bladder cancer, liver fibrosis (Richter, 2003), as well as heightened risk of HIV transmission (Mbabazi et al., 2011). The highest parasite burden is usually borne by school-age children and the disease has been linked to anemia, stunting and cognitive deficits, leading to poor school performance and higher drop-out rates (Ezeamama et al., 2018). Due to these

*E-mail: dr551@cam.ac.uk

Work leading to this paper was funded by the Swiss Network for International Studies (SNIS), in collaboration with the World Health Organization. The author is greatly indebted to Théophile Mande in Ouagadougou for his work in obtaining the survey datasets from the the Department of Sectoral Statistics at the Ministry of Agriculture and Hydrological Development of Burkina Faso. The author also thanks the Schistosomiasis Control Unit of the Ministry of Health of Burkina Faso for access to parasitological data. The full survey datasets are not publicly available: for replication purposes, all data and code used in the paper can be found at https://github.com/daniele-rinaldo/economic_impact_schisto. This paper is the sum of two separate works: the empirical part is titled "The Economic Impact of Schistosomiasis" and is co-authored with Javier Perez-Saez, Penelope Vounatsou and Jean-Louis Arcand, where the author is first author and is responsible for the entire econometric modeling and estimations. The entire theoretical framework is single-authored, and so are all errors. The author declares no further competing interests.

life-long impacts, schistosomiasis exerts large health, social and financial burdens on infected individuals and households (King, 2010).

Water resources development aimed at alleviating poverty in rural schistosomiasis-endemic communities in sub-Saharan Africa has been shown to exacerbate disease transmission (Steinmann et al., 2006). Dams and irrigation schemes expand the suitable habitat for the aquatic snails that serve as the intermediate hosts of schistosomes, and also increase the frequency and density of human-water contacts during which infection can occur (Diakité et al., 2017). This effect is particularly marked in water-constrained regions, possibly due to the concentration of contacts in the few available water points (Steinmann et al., 2006). In developing countries, the sector which one would naturally expect to be the most affected by the disease is agriculture, particularly in its subsistence form (De Janvry and Sadoulet, 2010, Christiaensen et al., 2011). This is because populations that rely heavily on agricultural production are the ones that are the most exposed to infection and ultimately suffer the highest disease burden. Schistosomiasis may thus induce "poverty traps". However, its net economic impact and the underlying tradeoffs between water resources development and public health concerns have yet to be rigorously quantified. General results concerning the relationship between diseases and economic development can be found in Acemoglu and Johnson (2007), Bloom et al. (2004), Bleakley (2007), Bleakley and Lange (2009) and Bleakley (2010). Despite extensive evidence concerning the long-term health effects of endemic diseases in developing countries, there have been few attempts to quantify their economic impact. This is particularly true of parasitic diseases such as schistosomiasis. Attempts have been sparse and either focused on specific mechanisms of small scale, albeit of great importance, or lacked results of significant strength (Foster et al., 1967, Baldwin and Weisbrod, 1974, Weisbrod et al., 1974, Audibert, 1986, Audibert and Etard, 1998, Miguel and Kremer, 2004, Audibert, 2010).

The empirical focus of this paper is on Burkina Faso, a country where schistosomiasis is endemic both in its intestinal and uro-genital forms (Poda et al., 2004, Lai et al., 2015), in order to establish a baseline case for what is potentially a large negative effect, and where there are less confounders generated by the interdependence of economic sectors found in more diversified economies. Burkina Faso is a low-income, landlocked Sub-Saharan country with very limited resources. In 2016, 80% of its 18 million inhabitants worked in the agricultural sector. Burkinabé agriculture is the main component of the country's economy, employing roughly 80% of the population, and is mostly of the subsistence variety, with low crop and livestock productivity and high levels of inefficiency (World Bank Results, 2017, Udry, 1996). Diversification in the sector is low, although increasing, with cotton being the most important cash crop. Large- and small-scale water resources development projects have been completed in the past 30 years to support agricultural activities and reduce climate vulnerability, which have however exacerbated the prevalence of malaria and schistosomiasis (Boelee et al., 2010). Mass drug administration campaigns, initiated by the Schistosomiasis Control Initiative in 2005, have been successful in reducing morbidity in most regions with a national mean prevalence of around 5% in school-age children in 2010 (Ouedraogo et al., 2016).

Figure 1 shows a schematic representation of the main empirical interests of this paper. The first mechanism that I want to study is the economic impact of schistosomiasis via the burden the disease exerts on agriculture. Because of the complexity of the disease dynamics and the profound interlinkages with a large set of socioeconomic and environmental variables (Gurarie and Seto, 2009), this identification requires the use of detailed agricultural and household datasets, as well as precise information on disease prevalence on

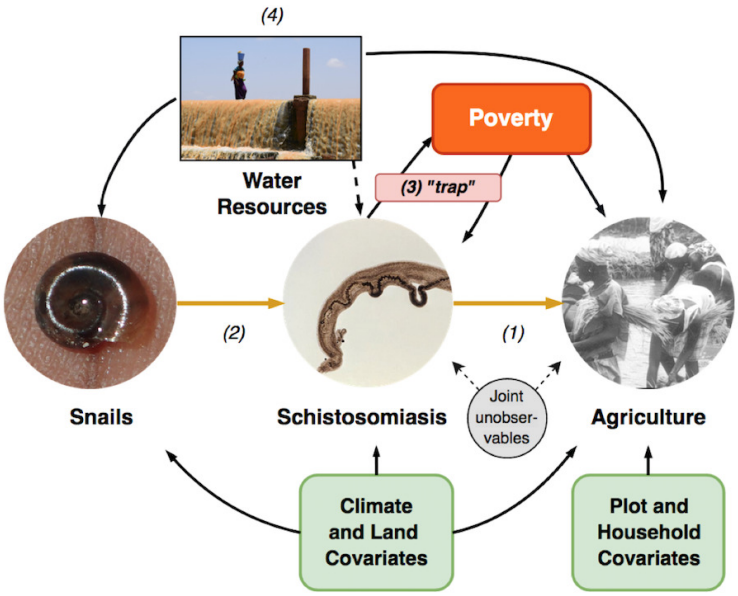

Figure 1: Mechanisms studied in the paper: (1) Effect of schistosomiasis on agriculture. (2) Freshwater snails, the disease's intermediate vector, as instrumental variables. (3) Poverty trap mechanics, where poverty and disease burden reinforce each other. (4) The feedback effects water resources development and disease.

a large spatial scale. Recent developments in disease mapping allow us to obtain high resolution prevalence maps which have been used in a variety of public health contexts (Lai et al., 2015) but have not hitherto been paired with household and plot-level data. I first create and geolocalize a large dataset for Burkina Faso that merges household and plot-level survey data with village-level disease prevalence data. I then proceed to estimate the effect of schistosomiasis on agricultural production with a variety of econometric techniques in the estimation of an agricultural production function. In order to address various estimation biases and establish causality of the effect, I use the spatial densities of the aquatic snails that serve as intermediate host of the disease as instrumental variables. In order to absorb all confounding effects of the large set of input variables and avoid having to assume a particular form for the production function, I exploit machine learning methods, and thus obtain a more precise estimate of the mechanism of interest. I estimate that the elimination of the disease would increase average crop yields by 7%, rising to 32% for the highest infection clusters, and show the disease burden to be akin to a productivity shock.

Having established the existence of an adverse economic burden of the disease, it is of great importance to study its consequences on poverty. Agricultural production in the countries where schistosomiasis is endemic, such as the sub-Saharan region and particularly Burkina Faso, is mostly of subsistence. This means that health shocks that translate to production shocks are likely to affect simultanously the income and the survival odds of farming households. I therefore focus on the lower tail of the joint distribution of overall crop size and crop weight farmed by households, which is a reliable indicator of household poverty, and find that in these tails the adverse effect of schistosomiasis is substantially greater. This shows that indeed poverty has a reinforcing effect on the burden the disease exerts on agricultural production, as well as being consequence of the burden itself: this loop is akin to a "poverty trap", creating states of indigence driven by the disease burden's debilitating nature.

The last mechanism I study empirically is the role of water resources development. I include in the dataset the a wide network of water infrastructures, ranging from the country's four main hydroelectric dams to small-scale irrigation reservoirs, and study the effect of their vicinity to villages on both agricultural yield and disease burden. The nature of this analysis requires the use of spatial techniques: the explicit inclusion of distance variables generates persistent spatial dependence in the data that cannot be removed even by the comprehensive set of fixed effects used in the rest of the paper. I show water resources infrastructures to considerably boost agriculture and economic development, but also to increase the adverse effects of schistosomiasis via both the increase of snail habitat and human-water contact. The presence of large-scale dams has substantial benefits on agricultural production, but simultaneously magnifies the negative burden of the disease on yield: villages located at a greater distance suffer a reduced burden. Returns to such large infrastructures are therefore significantly lower once their higher-order effects on health are explicitly taken into account.

The empirical results show schistosomiasis to generate a persistent negative productivity shock for farming households, reducing agricultural yield by its debilitating health burden and reinforcing poverty. I then construct a novel theoretical framework that uses these results as its structural justification, for example in the modeling of the farming households' agricultural production function. This allows the model to be based on assumptions firmly grounded on empirical evidence, framing it within a structure compatible with household behavior. This also allows the mechanisms I have uncovered in the data to be generalized beyond the specific case of Burkina Faso. I build on a rich existing literature on poverty traps, such as Matsuyama (1991), Azariadis and Drazen (1990) and Farmer and Benhabib (1994) up to Antoci et al. (2011).[1] The literature that joins disease dynamics to economic systems is more recent and more sparse: an example is Goenka et al. (2014), up to the recent flurry of research (Acemoglu et al. (2020) and McAdams (2020), among others). I join a standard household optimization problem with the dynamic equations of the disease established by the ecology literature. I show that the optimal choices of farming households facing the negative externalities due to schistosomiasis established by my empirical results can lead to equilibria where the disease is endemic and generates inferior levels of household wealth and consumption. I model a general human-schistosome contact parameter as a function of the distribution of wealth in the affected area, in order to represent the benefits stemming from increased overall sanitary conditions and better access to clean water, and to represent the reinforcing effect of poverty uncovered in the data.

I am able to obtain explicitly the global dynamics of the resulting system and show how even the simplest of frameworks, a basic household separable optimization problem, when augmented by the schistosomiasis dynamics it allows for the emergence of endemic equilibria characterized by high disease intensity and low household wealth and consumption. Such states are Nash equilibria, since schistosomiasis acts as an external constraint faced exogenously by the farming household, and is continuously determined by the individual wealth trajectories. I explicitly define the separator manifold between disease-free and endemic equilibria as the level of wealth for which the reproduction rate of the disease (the R_0, now wealth-dependent) crosses the threshold of unity. If the reproduction rate of the disease at these equilibria is high enough to dominate the positive externalities of wealth, even after a local perturbation of the fixed point, such Pareto-dominated states can mutate in absolute poverty traps from which the household cannot escape unless the initial joint

[1] For a comprehensive treatment of such literature we refer to Azariadis and Stachurski (2005)

state of wealth and disease intensity is drastically improved. In some cases, however, the inability of the farming households to escape such traps can derive from their incomplete information on the nature of schistosomiasis dynamics and especially on the feedback effects of wealth and agricultural production. Simply put, the farmers experience an observable level of schistosomiasis, but consider it only as an exogenously time-varying shock to their agricultural productivity and do not account for either its time evolution or the continuous feedback effects between wealth and disease intensity. This "information trap" therefore prevents them from the possibility of adjusting production decisions in order to escape such Pareto-inferior states of low wealth and consumption. This is because social optima, stemming from either fully informed households or a social planner explicitly taking the disease dynamics into account, require deviations from the separable optima that are proportional to the disease burden on the optimally controlled household utility paths.

Natural randomness in the environment can lead to fluctuations in the disease dynamics. For example, rainfall is largely responsible for creating the conditions that allow sufficient surface water accumulating in ponds and thus expand snail breeding sites; on the other hand, extended dry spells can drastically reduce snail habitat suitability. Natural variability derived from climate shocks, therefore, can directly affect the structural parameters that drive environmental and disease dynamics. I incorporate these realistic features in the model and study a stochastic system in order to better understand the seemingly erratic behavior that cannot be described by a deterministic framework. The stochastic extension of the model allows to capture the drastic shifts in the model structure caused by the omnipresent interdependences between the environment, diseases and human activity. I find that the interplay between wealth and stochastic disease dynamics, represented by the wealth-dependent reproduction rate, controls the probability flow between the disease-free and endemic states and acts as a regularizing force between the equilibria if the magnitude of the fluctuations is not excessive. I find that in the presence of environmental randomness the endemic state is characterized by a probability measure continuously modified by household decisions, with a stationary density shaped by the corresponding deterministic equilibrium of household wealth and consumption. This state of disease-driven poverty is a global attractor of the system, drawing wealth, consumption and disease trajectories towards it upon deviations from a disease-free condition.

The key policy implications of this paper are the following. First, elimination of the disease would boost both economic growth and farming households' well-being, and will help in drawing local communities away from persistent poverty states. The results further highlight the importance of shifting farming households towards more productive forms of agriculture, such as cash crops, as well as the necessity to increase awareness among the population of the anthropic factors affecting the dynamics of the disease. Another policy-relevant aspect of the results is the fact that smaller-scale water reservoirs, mostly used for irrigation purposes, do not seem to particularly affect the disease burden on agriculture, since I show that this feedback is generated entirely by large-scale dams, and therefore their presence in the territory should be incentivated.

This paper contributes to the literature in various ways, both from empirical and theoretical perspectives. The first one is the creation of a new dataset with the novel use of high-resolution disease prevalence maps together with a geolocalized survey dataset on farming households and crops. The second contribution is to establish the economic burden of schistosomiasis as substantial, poverty-reinforcing and akin to a productivity shock. The neglected state of schistosomiasis, and of its economic consequences in particular, was such that it has received far less academic attention than other diseases such as malaria, and my work reinforces

its relevance from a development perspective. The third contribution is the novel use of the spatial density of the disease's intermediate vector as instrument for the disease intensity, as well as using machine learning techniques to obtain a more precise estimate of the burden. The fourth contribution lies in the novelty of the theoretical model, founded structurally on the empirical results, which replicates the mechanisms found in the data and generalizes them outside our dataset. The model sheds further light on the nature of endemic states as Nash equilibria, both deterministic and stochastic, and how the observed poverty traps are a consequence of the wealth-dependent disease reproduction rate dominating over the positive feedback effect of wealth and being robust to localized shocks. The model has the attractive features of being very flexible, allowing for an explicit description of its rich dynamics and forming a baseline for multiple potential extensions and improvements, which I leave to future research. The model further highlights the fundamental importance in the fight against persistent poverty that lies in agents having complete information on the nature of the externalities they face, particularly in developing countries where natural constraints such as endemic diseases constantly strain economic systems and exhibit continuous feedback effects. Overall, my work reinforces the importance of accounting for the interactions between the environment, disease dynamics and economic development. While the empirical focus has been on Burkina Faso, in part because it is likely to be a worst-case scenario due to the country's joint economic and epidemiological profile, my framework can be applied to any country in which schistosomiasis is endemic and, indeed, to the economic impact of many other diseases.

The paper is organized as follows. Section 2 presents the dataset; section 3 presents the estimates of the burden of schistosomiasis on agriculture as well as the econometric framework and justifies the use of the instruments. Section 4 presents evidence of the disease-driven poverty trap and section 5 presents the results on water resources. Section 6 presents the theoretical model, its stochastic extension and the analysis of social optima. Section 7 gives an overall view of the paper and concludes.[2]

2 Data

Survey data were obtained from the National Institute of Statistics and Demography (INSD) in Ouagadougou. The annual plot-level agricultural dataset provides detailed information on crops, yields, inputs and pesticides, plot characteristics and labor type for the 2003-2017 period. Surveyed villages were distributed relatively uniformly across the country (Fig. 1a). The household survey data cover the 1996-2017 period and include detailed information on household characteristics and demographics. The agricultural dataset is at plot level while many of the variables in the household survey are at individual level. The main estimations of the paper are done at a plot level, therefore I have chosen to maintain all the information at plot level and aggregate individual variables in the household-level dataset at a household level, to then merge together via unique household identifiers into a large plot-level dataset. Villages are then geolocalized by matching the survey datasets with administrative level data by fuzzy string matching in order to bypass inconsistencies in village names due to different dialectal inflections and spelling variations. Households are not identified geographically, and I am able to localize the data only at a village level. I include a a comprehensive set of climatic remote-sensing data including precipitation, temperature, and vegetation indices. We extrapolate a variety of covariates from the raw rasters, such as mean temperatures in dry and rainy

[2]The appendix provides a full description of the data, provides a plethora of robustness checks for the empirical results, as well as all proofs.

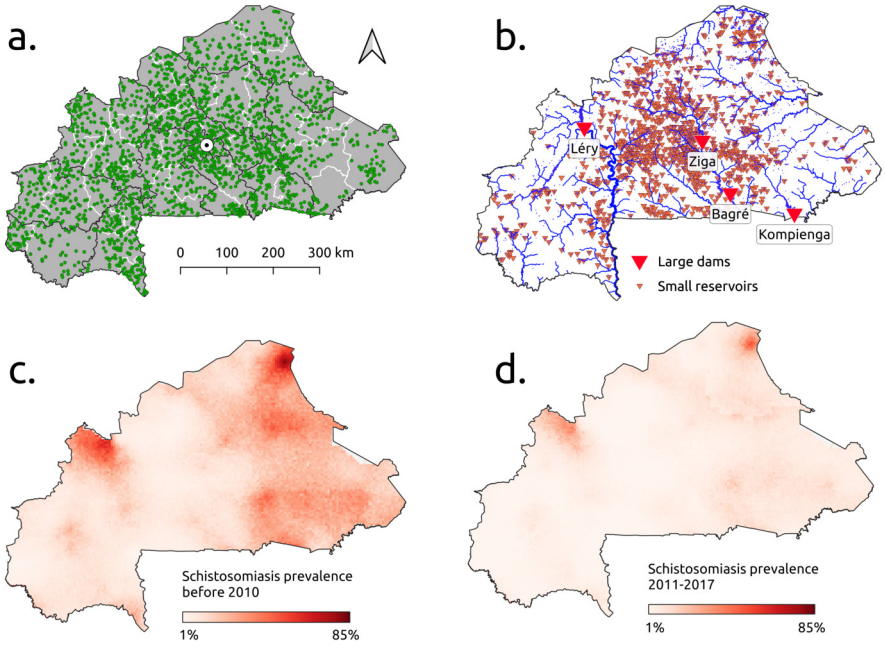

Figure 2: Data overview. (a) Villages included in the agricultural surveys, the capital Ouagadougou (white point) and level 1 (regions, black lines) and level 2 (provinces, white lines) administrative subdivisions. (b) River network (blue lines, width proportional to upstream area) and water resources infrastructure in the country. (c) Estimated schistosomiasis prevalence up to 2010. (d) Estimated schistosomiasis prevalence for 2011-2017.

seasons, precipitations in all seasons and mean and maximum dry spells. The pixel resolution of the climate data allows to merge the rasters with the survey datasets at the village level. I then include in the dataset the entire river and water resources infrastructure map for Burkina Faso, as shown in Fig.2b. I include the four main dams, shown with a label in the figure, as well as a variety of water reservoirs of both permanent and temporary regime, with covered surface ranging from around 1500 square meters up to 160 square kilometers. A comprehensive mapping summarizing the dataset is shown in Figure 2. Full information on its characteristics, summary statistics and details concerning the covariates used in the analysis can be found in Appendix A.1.

I then merge the dataset with two high-resolution maps of schistosomiasis prevalence estimates in school-aged children at a pixel resolution of 5×5 km. This combined dataset is one of the major contributions of this paper. The first map applies up until 2010 (Fig. 1c), the second thereafter (Fig. 1d). The estimated prevalence is a joint measure of both uro-genital and intestinal schistosomiasis, caused by *S. haematobium* and *S. mansoni*, respectively, and is obtained by means of Bayesian geostatistical analysis. For details concerning the techniques utilized in the creation of the prevalence maps, I refer to Lai et al. (2015) for all details of the first map, which apply equivalently for the second one. Because of resolution of the disease maps, prevalence is constant within each village: households belonging to the same village will be assigned the same level of schistosomiasis prevalence. Given that health consequences of schistosomiasis are more directly linked to infection intensity (measured by egg-output) than to prevalence (Audibert, 1986, King and Dangerfield-Cha, 2008), we translate estimated prevalence into a joint measure of schistosomiasis infection intensity in terms of the average number schistosome egg-output per person. This was done by assuming a negative binomial distribution of egg counts in urine and stools fit to parasitological data (see Appendix B.1 for details). All our results continue to hold when disease prevalence obtained directly from the maps is used in place of our intensity measure, as shown in Appendix B.3. Model-based estimates of spatial snail density were derived from malacological surveys collected in two field sites located along the South-North climatic gradient between the Sudanian and Sahelian regions (Perez-Saez et al., 2016).

3 Schistosomiasis and Agriculture

In order to estimate the economic impact of schistosomiasis, I rely on a specification that fully exploits the wealth of information available in our dataset at both plot and household levels. I identify the burden of the disease as a shock to overall productivity in an agricultural production function, which does not directly affect the amount of labor or physical inputs needed for production but instead influences their effectiveness. The production technology is specified as

$$Y_{ihjt} = A_{hj}(1 + m_{hjt})^{-\theta} F(X_{ihjt}) e_{ihjt}, \tag{1}$$

where Y_{ihjt} is the yield (output per hectare) of plot i, farmed by household h in village j at time t, m_{hjt}, which is common to all plots cultivated by a given household, represents the direct effect of schistosomiasis on productivity, calibrated by a parameter θ. I model the impact of the disease as $(1 + m)^{-\theta}$ in order to represent the fact that at a disease-free equilibrium the household's production technology is unaffected, and decreases as the disease burden worsens. The term A_{jh} represents household time-invariant productivity shifters, some of which may be unobservable, $F(.)$ is the production function and X_{ihjt} is the matrix of

total inputs, including plot, household, climate and land covariates. Some inputs vary only across village (climate), some are household-specific (all demographics and characteristics of the farming household) and some exploit the full plot-level variation, such as crop choice, plot characteristics and land quality. Lastly, $e_{ihjt} = \exp(\alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt})$ represents plot-level unobservables, which we decompose into crop- (c), household- (h) and time-specific components (t) as well as an idiosyncratic Gaussian disturbance term ϵ_{ihjt}. The parameter of interest is θ: it modulates the extent to which the disease affects agricultural yield, and its estimate is expected to be negative. Since the "real" effect ϕ of the disease is unobservable, I proxy it with the measure of infection intensity in terms of mean egg-output per person discussed earlier, denoted by $I_{jt} = (1 + Int_{jt})$. Taking logarithms then results in the quasilinear model:

$$y_{ihjt} = \tilde{A}_{jh} - \theta \tilde{I}_{jt} + \tilde{F}(X_{ihjt}) + \alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt}, \tag{2}$$

where the goal is to identify the parameter θ. In all estimations, for the variables that include zeroes (particularly plot surface and yield) I use the inverse hyperbolic sine transformation instead of the natural logarithm. I begin by imposing a Cobb-Douglas functional form on $F(.)$, which yields a fully linear model. The instrumented model takes the same form as (2). Note, if the household is assumed to maximize profits over its set cultivated plots, that $\tilde{F}(X_{ihjt}, K_{ihjt})$ must be replaced by $\tilde{F}(X^*_{ihjt}, K_{ihjt})$, where $X^*_{ihjt}(A_{hj}, I^{-\theta}_{jt}, p_{jt}, K_{ihjt}, \epsilon_{ihjt})$ is input use at the optimum, p_{jt} represents input prices and K_{ihjt} represent all other factors of production not under the control of the household (for example, climate). In this case the total marginal effect of schistosomiasis on yield would be given by $-\theta + \sum_{w=1}^{w=W} \frac{\partial \tilde{F}(X^*_w, K)}{\partial X_w} \frac{dX^*_w}{dI}$, where W is the total number of inputs chosen optimally. In Appendix B.2 I show that (1) is compatible with a model in which schistosomiasis affects *effective* labor input and that we estimate $\frac{dX^*_w}{dI} = 0$ for all $w = 1 \dots W$: schistosomiasis therefore is a pure productivity shock that does not affect optimal input use.

I then relax the restrictive assumptions (such as unitary elasticity of substitution between factors) imposed by the Cobb-Douglas functional form, and use adaptive machine learning methods in order to partial out all confounding and nonlinear effects of the factor inputs X_{ihjt}. The quasilinear structure of (2), stemming from the identification of schistosomiasis as a productivity shock, allows to disentangle the effect of the disease from the effect of the other covariates. After obtaining a consistent estimate of the function $\tilde{F}(.)$, one can partial out its predicted values from both log-yield and schistosomiasis infection intensity, thereby identifying the effect of interest, θ. Since determining the precise parametric form of $\tilde{F}$ that best fits the data is of secondary interest in the present context, adaptive machine learning methods are used, which are obscure in terms of interpretation but well-suited for prediction and classification. This procedure is similar to the procedure first proposed by Belloni and Chernozhukov (2013) and further expanded in Belloni et al. (2017) and Chernozhukov et al. (2018). The original procedure was proposed in order to deal with overfitting due to high-dimensional data and the estimation of treatment effects, and in this paper I extend it to the regularization of all confounding effects due to the unknown functional form for the production function from both outcome variable (yield) and disease intensity, after partialling out the required fixed effects: this extension is another methodological contributions of the paper. The procedure needs to be cross-fit to remove bias. The method requires imposing the double Neyman-orthogonal moment conditions:

$$\mathbb{E}[\psi(W, \theta, \eta_0)] = 0, \tag{3}$$

$$\partial_\eta \mathbb{E}[\psi(W, \theta, \eta)]_{\eta=\eta_0} = 0, \tag{4}$$

where $\psi(W, \theta, \eta_0) = ((y - \mathbb{E}[y|X]) - (I - \mathbb{E}[I|X])\theta)(I - \mathbb{E}[I|X])$ is the Neyman-orthogonal score function, $\eta_0 = (\mathbb{E}[y|X], \mathbb{E}[I|X])$, and where X is the complete matrix of covariates; ∂_η is a functional derivative operator, and the second condition imposes validity of the estimators under possible deviations from η_0. For the estimation of the unknown nuisance functions $\mathbb{E}[y|X]$ and $\mathbb{E}[I|X]$, I use random forests, gradient boosting machines and neural networks to obtain their predicted values and identify the parameter of interest θ, and compare their performance using a mean-squared error criterion. For further references on the estimators, see Hastie et al. (2001), and for all details on the architecture see Appendix B.3. Unbiasedness of the estimator is ensured by cross-validation. This approach is easily extended to an IV procedure, under the maintained hypothesis that the exclusion restriction for the IVs hold, and adds one stage per instrument to the usual procedure in which the Neyman-orthogonal moment conditions are imposed on each of the IVs.

I further estimate a semi-parametric model in order to expose potential non-linearities in the effect of the disease, where θI_{jt} is replaced by the smooth function $\phi(I_{jt})$ in Eq (2). Here ϕ is estimated by minimizing the squared residuals with a smoothing penalty on the second derivative of ϕ, instrumented using a standard two-stage semiparametric method. The smoothing parameter is chosen by cross-validation, and the degrees of freedom of the interpolating function are chosen iteratively by checking for zero signal on the residuals.

3.1 Snail densities as instrumental variables

The granularity of the schistosomiasis prevalence maps results in a disease measure that is constant at a village level. Estimating the mechanisms of interest, particularly the effect of schistosomiasis on agricultural yield and the characterization of the disease as a poverty trap, greatly benefits from the use of our rich dataset at a more disaggregated level, such as household or plot level. However, this is likely to suffer from a variety of endogeneity issues, especially related to error in measurement given that the "real" value of the disease burden is likely to vary within each village. We therefore address these issues with the use of snail densities as an IV. This approach requires a detailed justification. The prevalence measure that I use in the analysis is a joint measure, generated by averaging the two forms (intestinal and uro-genital) and subtracting the covariance, as in Lai et al. (2015). One therefore needs to instrument the two forms of the disease by including information on the different species of snail hosts. The intestinal form caused by *S. mansoni* is concentrated in the southwest part of the country, and its range is constrained by the presence of its intermediate host snails, *Biomphalaria pfeifferi*, as shown by Perez-Saez et al. (2017). This snail species is not present outside of this region due to its sensitivity to prolonged habitat dryouts which are more common in the Central and Northern parts of the country (Poda et al., 2004, Perez-Saez et al., 2017). As an instrument for intestinal schistosomiasis we therefore use the mean of *Biomphalaria* abundance for the villages in the southwest. The uro-genital form of the disease, caused by a different species of schistosomes, *S. haematobium*, is spread more uniformly throughout the country due to the ubiquity of the snail species of the genus *Bulinus* which serve as intermediate hosts. The highest prevalence of uro-genital schistosomiasis is found in the northern part of the Sahel. As opposed to *Biomphalaria*, *Bulinus* are present in a wide range of natural and man-made habitats throughout the country Poda et al. (2004), and we use estimates of its abundance in ponds and zones with ephemeral rivers in both the rainy and the dry season in order to capture changes in river size. I use the gridded predictions of the seasonal variations of snail densities produced in Perez-Saez et al. (2019). Joint dependence of snail density, schistosomiasis prevalence and agricultural yield on climate variables is accounted for by including a large set of climate covariates in all regressions, as explained in the Data section of the paper and in Appendix A.1. The main driver of snail predicted density is autocorrelation, as well as

a lagged structure of climate variables: results are unchanged if one uses for each disease year the lagged densities, in order to check for possible joint dependence on environmental unobservables for yield and snails. The concern was minimal, given the large range of environmental and climate covariates included, and the results confirm this. Error correction due to generated instruments is addressed by including model error in the estimations. In the dataset, the correlation between disease intensity and snail abundance is positive and strong, especially with *Biomphalaria* in rivers in the southwest, and with *Bulinus* in the northern dry regions where rivers are more ephemeral. In Appendix B.3 I report all tables showing relevance of the instruments and a control function approach implying rejection of the hypothesis of exogeneity.

3.2 Results

I present results of the estimation of (2) for households and plots observed in 2009 and 2011. These years are chosen to maximize the number of observations and provide the best fit and associated model diagnostics. Furthermore, these years cover the sharp decrease in disease burden in school-aged children through Burkina Faso's schistosomiasis control program observed between 2008 and 2013 (Ouedraogo et al., 2016). All results instrument schistosomiasis intensity with our snail density measures. Results are similar for other combinations of years for which there are enough repeated households. All estimates include a measure of malaria prevalence in order to account for co-morbidity which, interestingly, does not significantly affect yield at conventional levels of confidence. In no way does this imply that malaria has no adverse effect on households: this simply reflects the fact that malaria exerts a burden of an entirely different nature, possibly more compatible with a shock to labor supply. All additional tables, results for both plot- and village-level estimations and robustness checks are reported in Appendix B.3. The upper panel of Figure 16 reports the main estimates of the paper, which quantify the loss of agricultural yield due to schistosomiasis. The point estimate represents the marginal percentage impact of one additional worm egg per person on yield, and the labels indicate the percentage loss of agricultural yield due to schistosomiasis, both on average and in the top 5% infection intensity clusters.

The topmost estimate in the upper panel of Figure 16 results from a multiplicative (Cobb-Douglas) form for the production technology, which results in a log-linear model. The point estimate shows that schistosomiasis causes an average loss of agricultural yield of 8.9% (95% CI: 1.1%-16.5%), rising to 43.5% for the households in the top 5% quantile of disease intensity. The second estimate is agnostic concerning the functional form of the production technology: all the non-linear confounding effects stemming from the large matrix of inputs are absorbed by means of different adaptive machine learning methods, compared with a mean squared error criterion. By doing so, point estimates of the the disease intensity's marginal effect fall, and precision improves. For our 2009-2011 sample, our preferred method is given by tuned random forests, and results in a mean loss of yield caused by schistosomiasis of 6.6% (95% CI: 2.2%-12.02%), increasing to 32.2% for the household at the top 5% infection quantile. All plot-level estimates control for time, household and crop unobservables, and in Appendix B.3 we show how the endogeneity bias is likely to result from measurement errors, since estimates using heteroskedasticity as internal instruments (Lewbel, 2012) yield almost identical estimates. Aggregating the data to the village level yields similar results. Non-linearities in the impact of schistosomiasis are significant, as illustrated in the lower panel of Fig.16. The adverse effect of the disease is concentrated in villages that experience mid- to high levels of disease intensity, with the negative effect on yield appearing at intensities above 30 worm eggs/person. The upshot from the policy perspective is that the adverse impact of schistosomiasis on economic development is substantial, and control

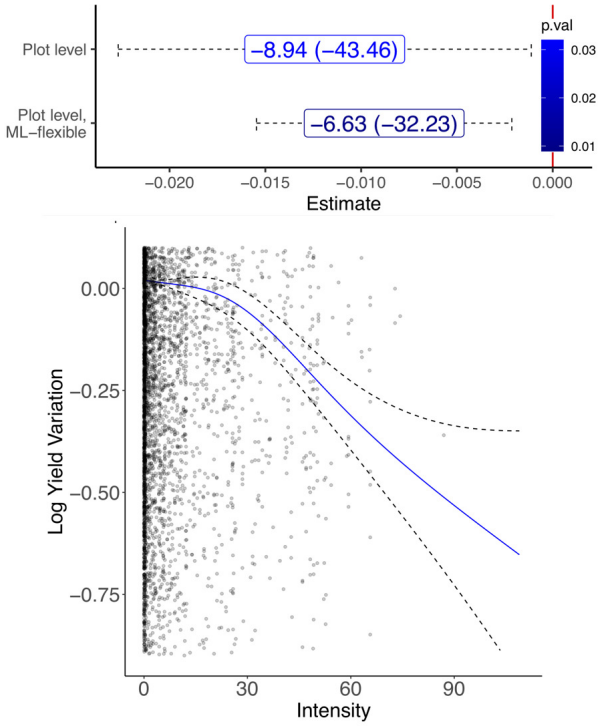

Figure 3: (Upper panel) Estimates of the yield loss (in percent) due to schistosomiasis for log-linear estimations (above) and with a flexible specification using adaptive ML methods (below). The point estimates indicate the marginal yield loss due to an extra worm egg/person. Each label reports the average loss and in parentheses the loss at the top 5% infection intensity clusters. 95% error bands are cluster-bootstrapped at the village level. (Lower panel) Nonlinear effect of the disease. Estimation done by fitting an instrumented adaptive spline using a two-stage semiparametric method. Dotted lines represent the 95% confidence interval in function fitting.

efforts aiming at reducing disease morbidity would produce substantial gains in agricultural productivity.

4 Schistosomiasis and poverty traps

In Burkina Faso, agricultural households are largely engaged in subsistence farming: shocks to yield therefore affect simultanously both income and survival probability. Moreover, it is likely that the productivity decrease attributable to schistosomiasis is a function of various household characteristics. Here, I explore two such characteristics: cropping patterns and poverty. I first focus on plots that produce the main cash crop farmed by Burkinabé households, cotton. In Appendix C I show that such plots are significantly larger and seem to be immune from any deleterious effects of schistosomiasis. A potential reason for this could be that households farming cash crops such as cotton on large plots are on average much richer: they may therefore have readier access to clean running water and enjoy better sanitary conditions. In contrast, households that rely on food crops farm smaller plots and suffer the most from the productivity shock. This is a first sign of how schistosomiasis is effectively a disease of poverty, acting both as its cause and consequence, as indicated in relationship (3) in Figure 1. However, cash crops, particularly GMO-resistant varieties of cotton, are also sometimes farmed by poorer households in order to reduce the risk of adverse shocks. To account for this, I examine the additional burden of the disease for households farming plots that belong to different quantiles of the joint distribution of plot surface and crop weight. Households in the lower reaches of this joint distribution correspond to those who are the most affected by poverty, almost entirely dependent on subsistence agriculture. I then interact schistosomiasis intensity with an indicator representing whether a given plot belongs to a specific quantile of the joint plot surface/crop weight distribution. The estimated equation is the following:

$$\begin{aligned} y_{ihjt} &= \tilde{A}_{jh} + \theta \tilde{I}_{jt} + \theta_{pj} \tilde{I}_{jt} \times \mathbb{1}_{(w_j, s_j) \leq Q_k^j} + \tilde{F}(X_{ihjt}) + \alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt} \\ Q_k^j &= \min\{w_{ihjt}, \mathrm{s}_{ihjt} : k \leq \Phi(w, \mathrm{surf}; t)\}, \end{aligned} \tag{5}$$

where Φ is the joint distribution of plot weight w and surface s for each year t. We estimate θ_p, the coefficient associated with this interaction variable, while varying k on a grid ranging from the 20th to the 60th percentile of the joint distribution. The coefficients associated with these variables represent the added burden of schistosomiasis linked with the underlying plot characteristics, which are likely to be correlated with poverty at the household level. In the upper panel of Figure 4 I show such plots suffer an additional loss to yield due to schistosomiasis ranging from 5% to 10%. These incremental effects vanish at the 55th percentile (the complete set of results and all details on coefficients and their standard errors is presented in Appendix D).

In order to characterize this mechanism in terms of its link to household poverty, I then carry out a similar procedure where the indicator function is defined at the household rather than at the plot level. The interest is in the estimation of θ_{ph} in

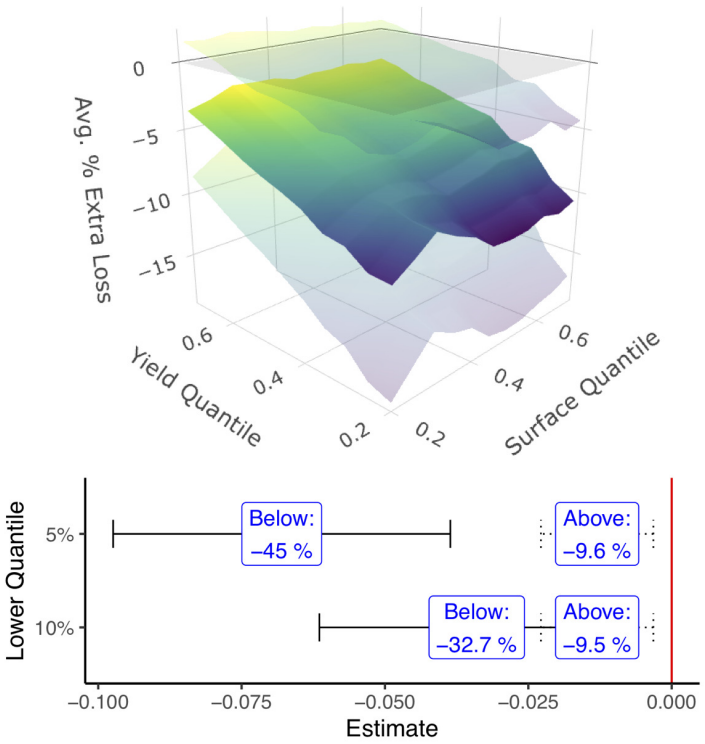

Figure 4: (Upper panel) Added schistosomiasis-induced yield loss due to poverty. Each point on the middle surface represents the extra loss due to the disease for plots below the respective crop weight and surface quantiles. The upper and lower transparent surfaces are cluster-bootstrapped 95% confidence intervals. (Lower panel) Losses to yield suffered by households above and below threshold levels of poverty, defined by the left tail of the joint harvest weight and plot surface distribution.

$$
\begin{aligned}
y_{ihjt} &= \tilde{A}_{jh} + \theta \tilde{I}_{jt} + \theta_{ph} \tilde{I}_{jt} \times \mathbb{1}_{(w^h, s^h) \leq Q_k^h} + \tilde{F}(X_{ihjt}) + \alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt} \qquad (6) \\
Q_k^h &= \min \left\{ \sum_{j=1}^{N_t^h} w_{ihjt}, \sum_{j=1}^{N_t^h} \mathrm{s}_{ihjt} : k < \Phi\left(w^h, \mathrm{s}^h; t\right) \right\},
\end{aligned}
$$

where $\Phi\left(w^h, \mathrm{s}^h\right)$ is the joint distribution of total crop weight and plot surface farmed by each household and N_t^h is the number of plots farmed by each household each year. The cutoffs k in this case correspond to the lower 5% and 10% tails. The lower panel of Figure 4 shows that for such households schistosomiasis exerts a disproportionately higher burden: -32.7% for the households in the bottom 10% and -45% for those in the bottom 5%. Results are unchanged if one carries out the procedure using the joint distribution of harvest and plot yield. Estimates in both specifications are from a log-linear model where schistosomiasis intensity is instrumented with a control function approach, and errors are cluster-bootstrapped at a village level.

Poverty thus reinforces the negative economic impact of schistosomiasis, with this feedback loop potentially generating a poverty trap phenomenon. Conversely, development interventions that increase agricultural productivity and allow peasants to diversify into cash crops will both improve living standards and reduce the burden of the disease.

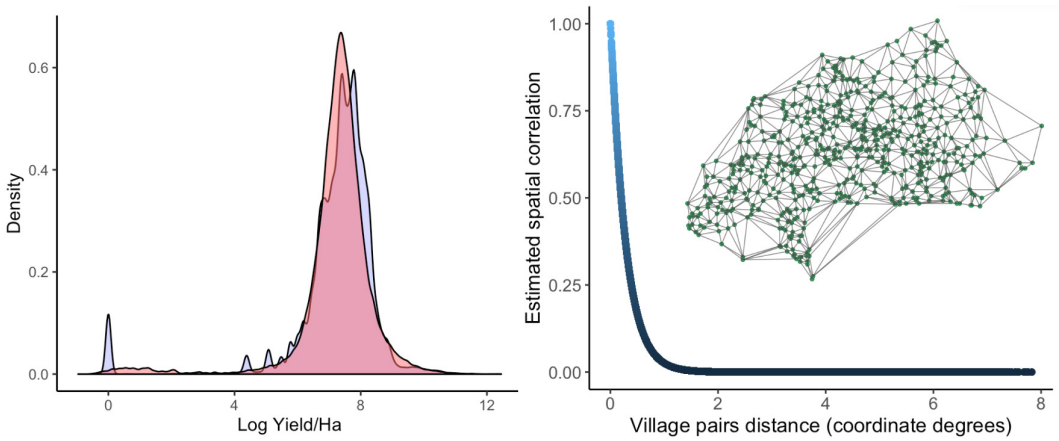

Figure 5: (Left panel) Homogeneity of households: original density of yield vs. rescaled density with village fixed effects partialled out. (Right panel) Distances between villages as spatial weights (inset), and the estimated spatial correlation.

5 Schistosomiasis and water resources development

Having established the reinforcing effects of schistosomiasis on poverty, it is of substantive importance to examine whether economic policies aimed at lifting people out of poverty can indirectly increase poverty itself by means of their impact on the spatial distribution of the disease, as indicated by mechanism (4) in Figure 1. Given that schistosomiasis is water-based, we study the feedback effects between the disease and water resources development. The first focus is on Burkina Faso's four main dams: the Bagré Dam and Kompienga Dam in Kompienga province, the Ziga Dam in Oubritenga province and the Léry Dam in Nayala province (Fig. 1b). To address this question I aggregate the data up to the village level, and compute the geographical distance of each village from the closest dams and reservoirs. An interesting characteristic of our dataset is the homogeneity of households in terms of yield: in Figure 5 I plot the density of yields in levels as well as the corresponding density after the within-village transformation. The densities are almost visually identical, and a Kolmogorov-Smirnov test cannot reject the null of equal distributions. A similar finding is presented by Udry (1996), using Burkina Faso ICRISAT survey data. Furthermore, the yield density remains equally unchanged after partialling out household unobservables: plot-level heterogeneity, therefore, seems to be driven by plot characteristics. Given the aforementioned homogeneity of household yields, village aggregation leads to what is essentially a representative household model.

The interest lies in establishing whether the presence of a large dam affects the magnitude of the effect of schistosomiasis on agricultural yields, and thus to identify villages in provinces that directly benefit from the presence of the four main dams. Interacting the presence of a dam with our measure of disease intensity allows to disentangle the direct impact of dams, which should increase yields, from the deleterious indirect effects that they may produce by facilitating the diffusion of schistosomiasis. The explicit inclusion of the dams, however, generates spatial dependence in the data that does not vanish even when controlling for the highest possible level of unobservables (regions), and therefore for the estimation we rely on a spatial autoregressive specification (SAR) of the form

$$y_{jt} = (1_{2j} - \rho W_j)^{-1}[\theta_1 I_{jt} + \theta_2 \mathbb{1}_{dam} + \theta_{int} I_{jt} \times \mathbb{1}_{dam} + \tilde{X}_{jt}\beta + \alpha_t + \epsilon_{jt}], \tag{7}$$

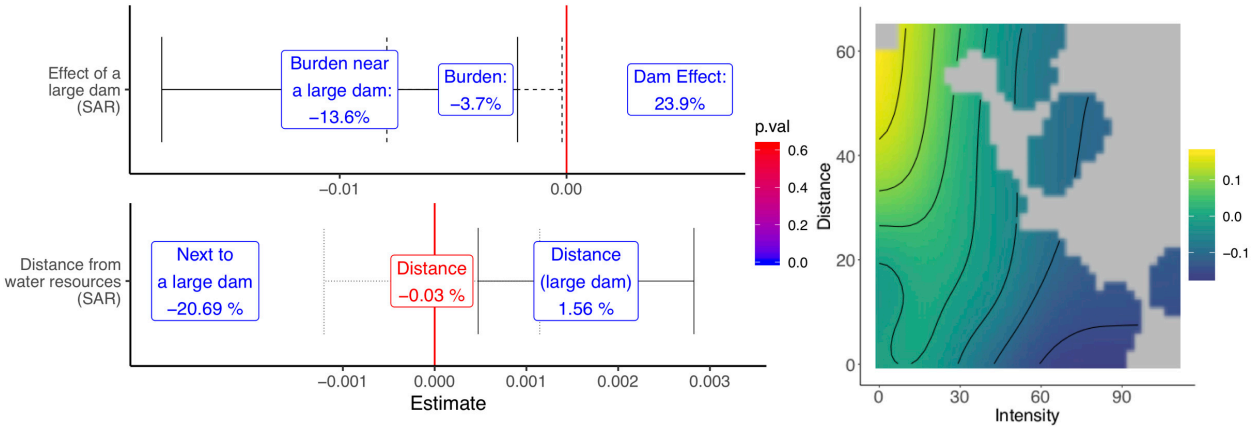

Figure 6: (Top left) Added effect of schistosomiasis on yields caused by the presence of a large dam. (Bottom left) Joint effect of schistosomiasis, distance from water resources networks and dam size. Estimations account for spatial correlation. (Right) Joint effect of schistosomiasis and distance from dams and water networks. Each point in the fitted surface represents the effect of schistosomiasis on yield for a village at the corresponding distance from a dam or a water reservoir: the darker the color, the more negative is the effect.

where W_j is a matrix of spatial weights given by the between-villages distance in coordinate degrees and ρ is the spatial correlation parameter, fit by maximum likelihood. A visual representation of the network of village distances used as spatial weights is shown in the right panel of Figure 5. The panel also shows how spatial correlation decays quickly as geographical distance (in coordinate degrees) increases, when the presence of dams is not included in the analysis. This makes spatial analysis irrelevant for village-level regressions without the inclusion of the water resources variables, once the appropriate fixed effects are included (region, province or commune): in Appendix E I show how once dams are included, spatial correlation remains even after including the same fixed effects, and thus requires the explicit use of the spatial techniques. We report all details concerning this class of estimates and their relevance for this set of estimations, as well as a full set of results obtained with different specifications. The upper panel of Figure 6 shows the results of estimating (7), showing that whilst the presence of a dam increases yield by 23.9% (θ_2), the estimate of θ_{int} shows that the average burden due to schistosomiasis increases from a yield loss of 3.7% to 13.6% for villages in proximity to large dams.

The previous results are refined by accounting for each village's distance in km from the nearest dam or water reservoir. The full network of dams and reservoirs is illustrated in panel (b) of Figure 1. I estimate an adaptive spline for the interaction term $I_{jt} \times dist_j$, with time and region fixed effects. The right-hand panel of Figure 6 shows how the deleterious marginal effect of schistosomiasis intensity on yield is mitigated as one moves further away from a dam or a reservoir, as well as how areas with high schistosomiasis intensity are concentrated within 20 km of a dam or reservoir. Households located in these areas suffer from large negative feedback effects between schistosomiasis and water resources development; to make matters worse, the effect of an increase in distance on the marginal effect of schistosomiasis intensity is greater for villages which display lower disease intensity.

From a policy perspective, it is important to investigate whether it is only large-scale dams that are the culprits when it comes to the transmission of the disease, or whether smaller scale projects built for livestock

and small-scale irrigation can potentially generate as much of an adverse effect. I interact the presence of a large dam with the distance from any water infrastructure as well as with the measure of intensity, and estimate the equation

$$\begin{aligned} y_{jt} &= (1_{2j} - \rho W_j)^{-1}[\theta_1 I_{jt} + \theta_2 dist_j + \theta_3 \mathbb{1}_{dam} + \theta_{3int} I_{jt} \times dist_j \times \mathbb{1}_{dam} + \theta_{2int1} I_{jt} \times dist_j + \\ & \theta_{2int2} I_{jt} \times \mathbb{1}_{dam} + \theta_{2int3} dist_j \times \mathbb{1}_{dam} + \tilde{X}_{jt}\beta + \alpha_t + \epsilon_{jt}], \end{aligned} \tag{8}$$

Results are presented in the lower left-hand panel of Figure 6, and again in Appendix E I report the full set of results with alternative methods and specifications. The estimate of θ_{2int2} For villages located within 1 km from a large dam, the average estimated loss due to schistosomiasis is 20.7%. The estimate of the triple interaction coefficient θ_{3int} shows that an increase of one kilometer in distance from the dam generates an average 1.6% *reduction* of the burden of schistosomiasis on agricultural yield: being further away from large dams is therefore beneficial. The distance effect for villages *not* in proximity of large dams (θ_{2int1}), however, is insignificant at any level of confidence: such dams seem indeed to be the main culprits of the feedback effects. The consequences from the standpoint of poverty and inequality can be substantial: populations that gain the most from such large irrigation projects often do not correspond to those most exposed to their deleterious consequences in terms of health and productivity.

6 Theoretical Framework: Schistosomiasis Dynamics, Household Decisions and Poverty Traps

Let us now examine the joint dynamics of optimal household behavior and schistosomiasis transmission. The first step is to model the natural dynamics of the disease: we use the following dynamic specification, as proposed by Woolhouse (1991) and firmly established in the ecology literature:

$$\dot{m}_t = \alpha N s_t - \gamma m_t \tag{9}$$

$$\dot{s}_t = BHm_t(1 - s_t) - \beta s_t \tag{10}$$

In this specification, $m_t \in \mathbb{R}^+$ is the number of schistosomes per person, and $s_t \in [0, 1]$ is the prevalence of patent infection of snails. The evolution of m depends on the rate at which schistosomes are gained by the human population, $\alpha N s$ minus the rate γm at which schistosomes die, where α is the rate of infection with schistosomes per person per snail, N is the number of snails and γ is the per capita mortality rate. The evolution of snails depends on the rate if increase in prevalence, $BHm(1 - s)$, minus the rate of decrease in prevalence, βs, where B is the per capita rate of infection of snails per schistosome and β is the per capita mortality rate for infected snails. Finally, H is the total number of humans, making the term Hm the total number of schistosomes.

As a first analysis, one can reduce the dimensionality of the dynamic system and use the fact that $\beta >> \gamma$, since the rate of turnover of the infected snail population is normally much higher than the schistosome one, and can assume that the prevalence of the snail population is in equilibrium and express

schistosome population dynamics as function of m_t alone. This is a perfectly acceptable assumption, as shown by Woolhouse (1991). The model then reduces to

$$\begin{aligned} \dot{m}_t &= \alpha \frac{NBHm_t}{BHm_t + \beta} - \gamma m_t \\ &= \alpha g(m_t) - \gamma m_t, \end{aligned} \tag{11}$$

where the parameters have the same interpretation as before.

6.1 Disease dynamics as external constraint: endemic states as Nash equilibria

I present a simple model of dynamic household optimization where each household takes disease intensity m_t as exogenous and does not include explicitly its dynamics within its optimization process. This implies that the model will not describe the joint dynamics of a globally controlled system, as in the case of an omniscient social planner who considers the joint state of both economy and ecology, but rather the individual decisions of a farming household who allocates its resources based on the observation of the level of schistosomiasis intensity they suffer. The aim is to endogenize the infection rate with individual behavior and optimization, in order to represent feedback effects of poverty, related to general health and sanitation characteristics as well as access to clean water. I therefore augment the standard model given by (11) with an infection parameter dependent on household wealth, given by $\alpha(x)$, where $\alpha_x < 0$, $\alpha_{xx} \geq 0$.[3] One can make this function dependent on the moments of the wealth distribution $\mathcal{P}(x)$, such as the area-wide average level of wealth $\bar{x}$, depending on the geographic scale of the model (region, province or even village), or include the effect of variance or other dispersion measures. I also set $\alpha(0) = \alpha$, where the parameter α represents a "maximum" level of schistosome to human transmission if wealth is identically zero. One therefore has an "internal" system, comprised by the household's evolution of wealth and the consequent consumption dynamics, and an "external" system, where the disease dynamics come into play and continuously affect household decisions. This is because the real interest is in a framework that can help explaining the empirical findings of farming households experiencing persistent schistosomiasis-driven poverty states, rather than one that concerns global optimal choices. This modeling framework is simple but goes a surprisingly long way in explaining our empirical findings. Furthermore, an attractive feature stemming from its simplicity is that a full exploration of its properties is explictly available, as well as the uncovering of a wide range of possible local dynamics around the multiple fixed points. The explicit nature of the results allows for precise characterizations of the global system dynamics without having to resort to numerical analyses which, whilst of fundamental importance, depend crucially on parameter choices.

Let us assume a household agricultural production function given by $Y_t = F(x_t, m_t, l_t)$, where $Y_x, Y_l > 0, Y_{xx}, Y_{ll} < 0, Y_m < 0$, where the impact of schistosomiasis is compatible with a Hicks-neutral productivity shock calibrated by $\theta > 0$. As shown by the empirical results, a modeling choice that is compatible with the data is a function of the form

$$Y_t := F(x_t, m_t, l_t) = A(1 + m_t)^{-\theta} f(x_t, l_t, K), \tag{12}$$

[3]The convexity assumption is to represent the effect of wealth on the human-schistosome contact function having a large marginal effect in a state of poverty, decreasing as wealth increases. In any case, the assumption on the second derivative does not affect any conclusion of the model, and is therefore left as a modeling guide for numerical analyses.

although in all what follows we choose to keep the functional form completely general. The production technology $f(.)$ is also made dependent on a set K of n inputs K_i, $i \in 1, \ldots, n$, not subject to optimization, such as available non-purchased cropland and climate variables, as well as within-household available labor. Coherently with our empirical evidence, we include in the production function the entire set of inputs optimized separately, implying the equivalence between household optimization and profit maximization, and without interdependence between input allocation and schistosomiasis intensity. I posit a standard separable optimization problem in which a continuum of farming households optimizes with respect to consumption and hired labor, disease dynamics are not known by the household and the disease intensity m_t is taken as an exogenous time-varying production factor. One then has

$$\begin{aligned} \max_{c,l} \quad & \int_0^\infty e^{-\rho t} u(c_t) dt && (13) \\ \text{s.t.} \quad & \dot{x}_t = F(x_t, m_t, l_t) - \delta x_t - c_t - w l_t \\ & x_t \geq 0 \ \ \forall t \\ & [x_0, m_0] = [x^0, m^0] \in \mathbb{R}^+. && (14) \end{aligned}$$

where $u(.)$ is the concave household utility function, ρ is the discount factor, F is the production technology as given by (12), w is the exogenous wage and δ is the household wealth depreciation (in an agricultural setting, this can be interpreted as the harvest loss). As stated before, this implies that the household's consumption decisions and wealth are continuously affected by the presence of the disease, as it's directly affecting agricultural productivity and observed yield. However, the household does not account explicitly in its decisions for the disease dynamics and how their production outcome can affect its evolution, and simply considers the suffered level of schistosomiasis intensity as a time-varying parameter, akin to an exogenously determined price. The wealth-dependent schistosome-human contact function becomes therefore a positive externality of the optimization process of the farming households.

It's easily seen that labor is optimized in a static way and $l_t^* = F_l^{-1}(w)$ for all t, coherently with the separable nature of the problem which implies that the optimal labor allocation is equivalent to a profit-maximizing one[4]. At each time step the optimal hired labor supply is a continuous function of the state, and clearly $\partial l^*/\partial x > 0$, $\partial^2 l^*/\partial x^2 < 0$ and $\partial l^*/\partial m < 0$. We substitute in the wealth dynamics the optimized labor supply, depending now only on wealth and disease, and write $F(x_t, m_t) = F(x_t, m_t, l_t^*) - w l_t^*$. Assuming parameter sets such that the dynamic optimization allows for non-degenerate optima [5], this "new" production function retains the same characteristics as the original, namely $F_x > 0, F_{xx} < 0, F_m < 0$. Consumption is optimized in a dynamic way: by a standard procedure the dynamical system of interest is therefore given by the dynamics of the vector $\mathbf{z}_t = [x_t, m_t, c_t]$, comprised of two state variables and one internal forcing (the control). It can be written as

[4] In order to be consistent with our empirical results, we would have to set $\partial \log F(x, m, l)/\partial \log(1 + m) = -\theta$, since partial derivatives in factor input regressions were consistently estimated at zero. This implies that we would require $l_t^* = f_l^{-1}(w)$, thus imposing that the farming household maximizes profits without the presence of the disease, thus adding another degree of separability in the overall optimization process. This does not affect the nature of the solutions, since $F(x, m, f_l^{-1}(w)) - w f_l^{-1}(w)$ remains a function increasing and concave in x and decreasing in m for all admissible parameter sets, although the overall magnitude of the marginal effects would clearly change.

[5] Namely, that $\partial F(x, m)/\partial x > 0$, which holds using the estimates obtained in the agricultural regressions.

$$\begin{aligned} \dot{x}_t &= F(x_t, m_t) - \delta x_t - c_t \\ \dot{m}_t &= \alpha(\mathcal{P}(x))g(m_t) - \gamma m_t \\ \dot{c}_t &= -\frac{u'}{u''}\left[F_x(x_t, m_t) - (\delta + \rho)\right]. \end{aligned} \tag{15}$$

The system (15) is paired with a standard endpoint condition on consumption and co-state (limiting to zero in infinite time) allowing for non-explosive behavior. We are here considering a continuum of farming households who do not consider explicitly the disease dynamics, but their production output (wealth) affects the human-schistosome contact function. The resulting paths of the system (15) are Nash equilibria rather than socially optimal outcomes, deriving from the fact that the externality of wealth on disease $\alpha(\mathcal{P}(x))$ is the outcome of production decisions of identical agents without any incentive to deviate from the "internal" optimization process, where schistosomiasis is exogenously determined. We therefore have $\alpha(\mathcal{P}(x)) = \alpha(x_t)$, and density effects are equivalently represented by individual trajectories. Agents being identical is evidently a rather strong simplifying assumption, but it is also not rejected by the homogeneity of farming households observed in the data (see the right panel of Figure 5).

The disease dynamics (11) allow for two distinct disease equilibria: the first, is a disease-free one that generates a steady state $\mathbf{z}^*_{df} = \{x^*_{df}, 0, c^*_{df}\}$, in which the levels of wealth and consumption x^*_{df}, c^*_{df} are identical to the standard optimization problem without the presence of the disease. The second is the endemic equilibrium given by

$$m^*_e(x_t) = \frac{\alpha(x_t)N}{\gamma} - \frac{\beta}{BH} = m^*_e(x_t). \tag{16}$$

which depends on the level of household wealth x_t. Since $\alpha_x < 0$, the poorer the households are and the higher will this endemic level be. Associated to this endemic level of disease there will be an endemic steady state $\mathbf{z}^*_e = \{x^*_e, m^*_e, c^*_e\}$, with associated equilibrium levels of household wealth and consumption x^*_e, c^*_e, and the resulting endemic level $m^*_e(x^*_e)$. If $m^*_e(x^*_e) > 0$, then an endemic equilibrium is both possible and compatible with household wealth dynamics, and setting the time derivative of wealth and consumption in (15) to zero at m^*_e one obtains immediately that the endemic levels of wealth and consumption are strictly inferior to its disease-free counterpart, i.e. $x^*_e < x^*_{df}, c^*_e < c^*_{df}$. It is therefore a equilibrium Pareto-dominated by the disease-free one, with lower consumption and lower wealth, as well as a higher disease incidence implying reduced household health and overall welfare. The difference between the two equilibria depends continuously on the model parameters, particularly on the incidence of schistosomiasis in the production function θ, and we have that $dx^*_e/d\theta < 0$, as shown in Figure 7. If $m^*_e(x^*_e) < 0$, then no endemic equilibrium is possible, and the system allows only the disease-free steady state. The disease-free case only represents the case of all the regions and countries in which there is no schistosomiasis and an endemic state cannot be sustained: this can be a consequence of numerous factors, such as greater wealth, lower infection parameters or low suitability for snail habitat. For such systems, even if the disease was introduced by external contagion, the dynamics of the system eventually fade the disease intensity to zero in finite time. This case is of limited interest here, as we focus on the case where the disease can generate a persistent adverse burden: from now on we will only allow parameter sets that generate a positive endemic level of schistosomiasis, i.e. $m^*_e(x^*_e) > 0$, and study the characteristics of both the equilibria.

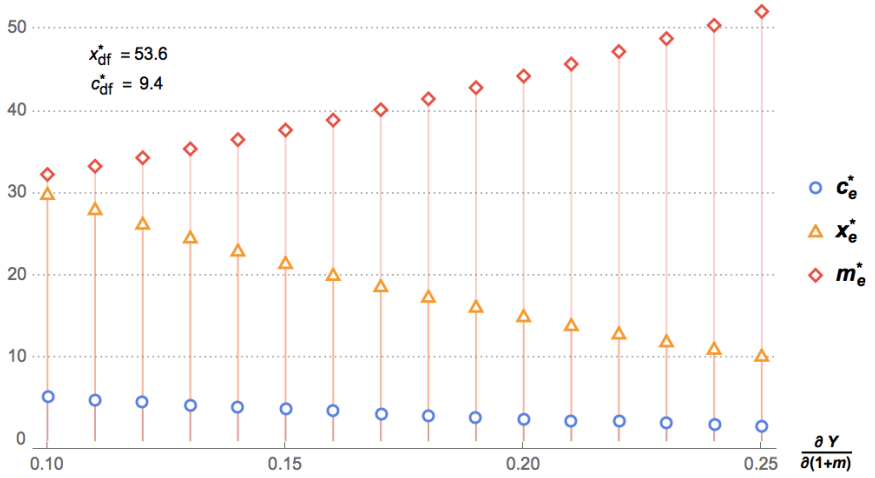

Figure 7: Pareto-dominated endemic equilibria with both wealth and optimal consumption lower than the disease-free counterparts x^*_{df}, c^*_{df}. As the marginal effect of the total burden of schistosomiasis $(1+m)$ on agricultural production (Y) increases, the associated endemic equilibria exhibit increasingly higher levels of disease intensity and lower levels of wealth and consumption. We model the wealth-dependent human-schistosome contact function as $\alpha(x) = \alpha/(1+x)^s$, with $s = 0.5$, and a Cobb-Douglas production function calibrated using the empirical estimates.

The determinant of the system's Jacobian evaluated at the disease-free equilibrium is given by

$$\det J(x^*_{df}, 0, c^*_{df}) = \left(\alpha(x^*_{df})g'(0) - \gamma\right)\left(-\frac{u'}{u''}(c^*_{df})\right)F_{xx}(x^*_{df}, 0).$$

Since $F_{xm} < 0$, one can immediately see that the term $\partial \dot{m}_t/\partial m_t$ plays a crucial role in determining the behavior of the system at the disease-free equilibrium: simplifying the expression one obtains that such quantity is positive if

$$\alpha(x^*_{df})\frac{BHN}{\beta\gamma} > 1. \tag{17}$$

The left-hand side quantity in (17) is known as the R_0 of the disease, or its reproduction rate: this inequality is the condition for endemicity for the disease, which controls whether asymptotically the disease will reach either the disease-free or the endemic equilibrium depending on whether this quantity is below or above unity. In our framework, this indicator is now wealth-dependent, and allows to explore the feedback effects of poverty on the disease diffusion and viceversa. If there were no such wealth effects and $\alpha(x) = \alpha$, then this quantity would simply be a scalar determined by the "natural" disease parameters, and if $R_0 < 1$ no endemic state would be possible, regardless of household choices.

The key role played by this quantity is seen by studying the eigenvalues of Jacobian evaluated at the disease-free equilibrium: they are given by

$$
\begin{aligned}
\lambda_1 &= \gamma\big(R_0(x^*_{df}) - 1\big) \\
\lambda_2 &= \frac{-\rho - \sqrt{\rho^2 + \left|-\frac{u'}{u''}(c^*_{df})F_{xx}(x^*_{df}, 0)\right|}}{2} \\
\lambda_3 &= \frac{-\rho + \sqrt{\rho^2 + \left|-\frac{u'}{u''}(c^*_{df})F_{xx}(x^*_{df}, 0)\right|}}{2},
\end{aligned}
$$

all three real and of which one is negative (λ_2) and one is positive (λ_3). This makes the disease-free equilibrium saddle-path stable, except for the case in which $R_0(x^*_{df}) = 1$ which makes the equilibrium non-hyperbolic. In this case stability cannot be determined *a priori*, and the local linearization may not be appropriate: however, R_0 determines whether schistosomiasis converges to zero or the endemic equilibrium, implying that the subspace $\{x, m, c\} \in \mathbb{R}^3$ such that $R_0(x) > 1$ around $\mathbf{z}^*_{df}$ generates the manifold tangent to the unstable subspace of J^*_{df}, associated to λ_3, and the subspace such that $R_0(x) < 1$ generates the manifold tangent to the stable subspace associated to λ_2. If $R_0(x^*_{df}) \neq 0$, then it determines the direction of the system trajectory upon perturbation based on whether it is greater than one or not. If $R_0(x^*_{df}) > 1$, the system is an unstable saddle with a two-dimensional unstable manifold since $\lambda_2 < 0 < \lambda_3 < \lambda_1$, implying that a perturbation of the initial disease-free state pushes the system away towards the endemic equilibrium unless the initial condition is on the only one-dimensional stable manifold. This result is intuitive: if even at the disease-free equilibrium the reproduction rate is greater than unity, the system is prone to disease endemicity and will tend towards the Pareto-dominated state. If $\lambda_1 < 0$, the disease-free equilibrium will be an attracting saddle, with only one unstable manifold corresponding to initial conditions generating a $R_0(x_0) > 1$.

It is of particular interest to study perturbations of this equilibrium: it's important to see whether shocks to either disease intensity, via its introduction via external contagion, or to wealth push the system towards the endemic state or ultimately allow the system to remain on the disease-free one. It's straightforward to show that the dominant eigenvalue (the one with the largest real part in absolute value) is λ_1, which means that the system is asymptotically pushed to move on the corresponding eigenvector $e_1 = (-sign(\lambda_1) \times |\vec{dx}|, \vec{dm}, -sign(\lambda_1) \times |\vec{dc}|)$. If λ_1 is positive, the eigenvector points towards a state with lower wealth and consumption and higher disease intensity since in dt we have $dm > 0$. Perturbations ϵ of the equilibrium also bring the system back immediately to the stable manifold and then to the disease-free equilibrium, if $R_0(x^*_{df} + \epsilon_1) < 1$ where ϵ_1 is the first element of the perturbation vector. This is better seen by using Morse's lemma for the direction of perturbed initial states:

$$
F(\mathbf{z} + \Delta\mathbf{z}) = F(\mathbf{z}) + \frac{1}{2}\sum_{i=1}^{\text{rank}(J)} \lambda_i \Delta\mathbf{v}_i^2,
$$

where $\Delta\ \mathbf{v}_i$ are the new parameters of the model corresponding to the motion of the system along the corresponding eigenvectors e_1. This means that the overall trajectory is determined by the dominant eigenvalue, which will push the perturbed system to a path along an eigenvector with positive values: wealth and disease intensity increase, consumption shifts to the stable manifold and wealth starts decreasing because of the increase in schistosomiasis. If the perturbation didn't push $R_0(x_{df} + \epsilon_1)$ beyond the endemicity threshold,

the increase in disease intensity eventually starts to diminish and reaches the disease-free equilibrium.

One now needs to study the characteristics of the endemic equilibrium, which allows for more involved dynamics. I can now show that there exist three possible scenarios, depending on the magnitude of the corresponding disease reproduction rate: two are saddle points, one repelling and another attracting, and the third is a fully attracting state (a trap). We begin by studying the determinant of the associated Jacobian, which is given by

$$\det J_e^* = \left(\frac{u'}{u''}(c_e^*)\right)\left[\alpha'(x_e^*)g(m_e^*)F_{xm}(x_e^*,m_e^*) - \left(\alpha(x_e^*)g'(m_e^*) - \gamma\right)F_{xx}(x_e^*,m_e^*)\right] \tag{18}$$

Since u'/u'' is negative, α is decreasing in wealth, F is increasing and concave in wealth and decreasing in m, a key condition for the negativity of the "endemic" Jacobian determinant is given by the sign of the term $\alpha(x_e^*)g'(m_e^*) - \gamma$. If this term is positive, the determinant is univoquely negative. Substituting in this term the endemic level of disease as a function of x_e^*, one can show the positivity condition to be equivalent to

$$\frac{\alpha(x_e^*)NBH}{\beta\gamma} = R_0(x_e^*) < 1, . \tag{19}$$

which is the wealth-dependent R_0 of the disease evaluated at the endemic equilibrium. Note that positivity of the term $\alpha(x_e^*)g'(m_e^*) - \gamma$ is now guaranteed if the R_0 at the endemic equilibrium is smaller than unity. This is the opposite of the disease-free case: it is intuitively so, since the nature of the two equilibria is connected to the respective $R_0(x)$. A repelling nature for the disease-free equilibrium is connected to a reproduction rate higher than unity: for the endemic equilibrium we have a similar condition. The condition is not anymore dependent on the unity threshold, since if $R_0(x_e^*) < 1$ the endemic equilibrium is not feasible. Furthermore, since $\partial R_0(x)/\partial x < 0$, if the endemic equilibrium is not feasible then certainly $R_0(x_{df}^*) < 1$, and the only possible equilibrium and the global attractor will therefore be given by the disease-free one given by the triplet $(x_{df}^*, 0, c_{df}^*)$, which exhibits saddle-path stability and one unstable manifold. So $R_0(x_e^*) > 1$: however, this condition is not sufficient for a description of the system dynamics around this equilibrium.

If $R_0(x_e^*) > 1$, then there exists a feasible Pareto-dominated equilibrium $\mathbf{z}_e^* = (x_e^*, m_e^*, c_e^*)$ with lower consumption and wealth and higher disease intensity. The nature of this equilibrium is slightly more involved than the disease-free case, since the eigenvalues can be obtained in closed form but the expressions are very large and not informative. We can however obtain analytical results on their sign by studying the form of the characteristic polynomial of the Jacobian at the endemic equilibrium: we have a left asymptote at ∞ and a right one at $-\infty$, and a $y-$intercept given by the determinant of the endemic Jacobian. It's easily shown that one of its inflection points has to be on the negative axis, therefore one eigenvalue has to be negative. The sign of the other two is therefore determined by the sign of $\det J_e^*$: if positive, another root of the characteristic polynomial has to be also negative, and if negative the system has either two positive eigenvalues or three negative ones[6]. The size of the endemic reproduction rate $R_0(x_e^*)$ is therefore again key in determining the nature of the Pareto-dominated equilibrium. Breaking condition (19) is however not sufficient for changing sign of the determinant at the endemic equilibrium: this equilibrium is non-hyperbolic if

[6]This is also implied by the determinant being the product of the eigenvalues.

$$\Delta R_0(x_e^*) := \frac{R_0(x_e^*) - 1}{R_0(x_e^*)} = \frac{1}{\gamma} \left| \frac{F_{xm}(x_e^*, m_e^*)}{F_{xx}(x_e^*, m_e^*)} \frac{\partial \dot{m}}{\partial x}(x_e^*, m_e^*) \right|, \tag{20}$$

which happens when the percentage deviation of the endemic R_0 from the threshold reaches a quantity equal to the curvature ratio of the production function inputs, modulated by the slope of the human-schistosome contact function, representing the tradeoff between the decreasing returns to scale of wealth and its decreasing effect on the disease diffusion.

If $\Delta R_0(x_e^*)$ is positive but smaller than the right-hand quantity in (20), then the endemic equilibrium is a *repelling saddle*, and the disease-free is attracting. This implies that most trajectories except for the single one-dimensional stable manifold will push away from endemicity and towards a disease-free state: this is a consequence of the R_0 being high enough to allow the existence of an endemic equilibrium, but not high enough to attract most trajectories to it. If $\Delta R_0(x_e^*)$ is greater than the quantity in (20), the equilibrium becomes attracting. There can still exist a manifold for which the system is pushed away, making the equilibrium an *attracting saddle*, but most initial conditions will attract towards it. This condition is based on the strength of the wealth-dependent reproduction rate to exceed an amount proportional to the interaction of household production characteristics and the positive externalities of wealth on disease diffusion. Note that the unstable manifold can lead both outside the $R_0 = 1$ surface and towards zero wealth and consumption. If the initial condition on wealth, schistosomiasis and consumption puts the household outside the part of the unstable manifold that allows it to escape the Pareto-dominated equilibrium and reach the disease-free state, then it will be trapped inside a state with high disease, low wealth and low consumption. The unstable manifold is however one-dimensional and unlikely to be reached by a random perturbation. This is a consequence of the household optimizing within an "internal" system, where schistosomiasis is exogenously given, but the actual wealth dynamics being determined by the "external" system, in which the disease dynamics are determined by wealth-driven Nash outcomes. It is furthermore driven by incomplete information: the household optimizes within a non-autonomous system, where the disease evolves exogenously in time and in which the household cannot be able to extrapolate which manifold would allow it to escape the poverty state, for example by modifying the labor choice or by a temporary modification of consumption.

There can be however a scenario in which the Pareto-dominated equilibrium is entirely unavoidable, and even if the household was aware of the external constraint generated by the disease dynamics. It can be shown that the Jacobian of the system (15) evaluated at the endemic state identifies a fully attracting equilibrium - a sink, where all three eigenvalues are negative - if the following conditions are satisfied, corresponding to the Routh-Hurwitz stability criterion[7]:

$$\Delta R_0(x_e^*) > \frac{\rho}{\gamma} \tag{21}$$

$$\frac{1}{\gamma} \left(F_m(x_e^*, m_e^*) \alpha'(x_e^*) g(m_e^*) + \frac{u'}{u''} F_{xx}(x_e^*, m_e^*) \right) < \rho \Delta R_0(x_e^*) \tag{22}$$

$$\left(\rho - \gamma \Delta R_0(x_e^*) \right) \left(\frac{\partial \dot{m}}{\partial x} \frac{\partial \dot{x}}{\partial m} + \frac{\partial \dot{x}}{\partial x} \frac{\partial \dot{m}}{\partial m} + \frac{\partial \dot{c}}{\partial x} \frac{\partial \dot{x}}{\partial c} \right) \Big|_{\mathbf{z}_e^*} + \det J_e^* > 0. \tag{23}$$

[7] See for example Takayama (1985) for an exposition of the related theorems. The result stems from straightforward algebric computations, albeit lengthy, of the characteristic polynomial of the Jacobian and the application of these conditions, and its proof is therefore omitted.

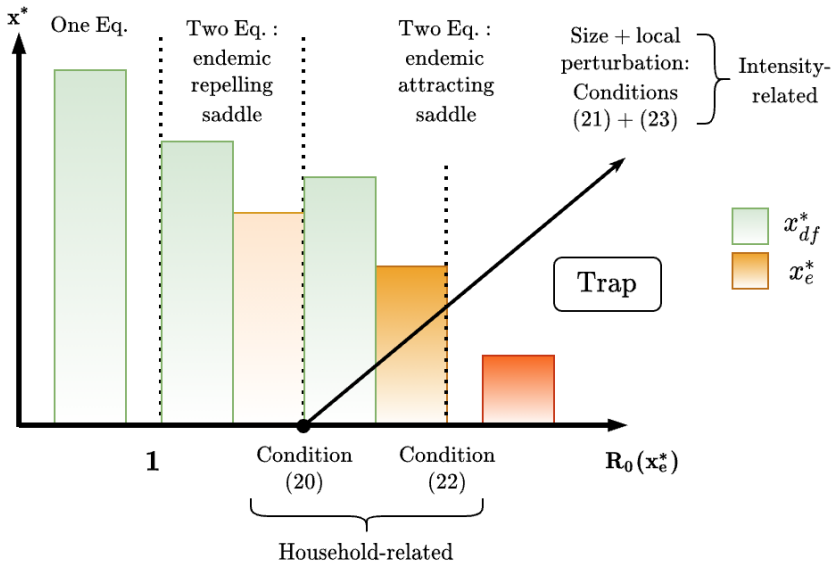

Figure 8: Schematic representation of the nature of the different equilibria. If the R_0 at the endemic equilibrium is less than unity, only a disease-free state is possible. If this rate dominates the household-driven conditions, the endemic equilibrium is an attracting saddle. If this quantity remains greater than unity after a local perturbation of the equilibrium, a poverty trap is created.

Condition (21) imposes that the proportional deviation of the R_0 at the endemic equilibrium from the unity threshold exceeds an amount equal to the ratio between the household's intertemporal discount rate and the schistosome death rate, and therefore is an "absolute" criterion on the magnitude of the disease's evolution rate. Condition (22) imposes a condition on the lifetime discounted endemic evolution rate. It has to exceed an amount that sums the marginal effect of the disease on the household decision system (the marginal effect of disease on production times the marginal effect of wealth on disease diffusion) with the marginal effect of wealth on optimal consumption. Note that the argument in parenthesis is always positive. This criterion compares disease impact magnitude with the household structure, and intuitively corresponds to a "harmful" part dominating a "beneficial" one, where the disease evolution weighed by its negative effects on productivity is strong enough to overweigh the reduction of disease intensity generated by higher wealth. Condition (23) extends the previous conditions and imposes another bound on the magnitude of the R_0 at the endemic equilibrium, weighed by the local system dynamics around the fixed point. It imposes that the magnitude of the disease reproduction rate at the endemic equilibrium, large enough to satisfy (21), needs to be large enough to keep dominating the "net" positive effects of wealth and to be robust to a local perturbation.[8] If the schistosomiasis reproduction rate at the endemic equilibrium satisfies conditions (21), (22) and (23), then this Pareto-dominated states is a sink, attracting all trajectories around, characterized by extremely low wealth and consumption and a high disease intensity. In this extreme case, even if a household was aware of the disease dynamics, unless its initial levels of wealth, disease and relative consumption allow for a decreasing level of m_t in dt, implying a negative $R_0(x_{dt})$, it will be drawn by the increasing levels of disease and consequent decreases in wealth towards this extreme state of poverty. The key role played by the joint dynamics of wealth and disease intensity in determining the nature of the equilibria, via the

[8]Note that condition (22) is satisfied if (20) holds, implying that $\det J_{\mathbf{z}^*_e} < 0$.

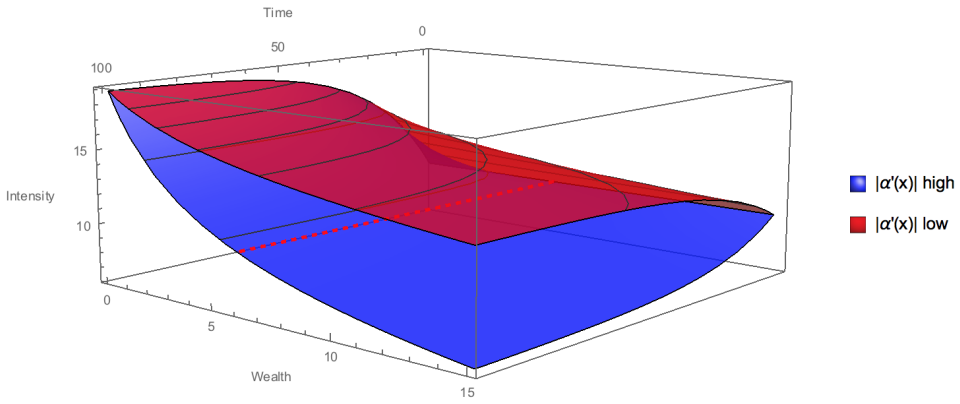

Figure 9: Disease trajectories for different levels of wealth. The red surface (above) shows that as wealth increases, the endemic level of schistosomiasis decreases but its corresponding equilibrium remains feasible. Here we model the wealth-dependent human-schistosome contact function as $\alpha(x) = \alpha/(1+x)^s$, with $0 < s < 1$. The blue surface (below) shows that if $\alpha'(x)$ is high enough, for a threshold level of wealth (dashed red line) the disease dynamics tend away from the endemic equilibrium and towards a disease-free state. The threshold level of wealth and the corresponding surface on $\{m, c\}$ is the separatrix for the basins of attraction of the two equilibria.

wealth-dependent disease reproduction rate, is summarized in Fig.8.

This hints at the nature of the wealth-dependent basic disease reproduction rate $R_0(x)$ as a *separatrix* on the $\{x, m, c\} \in \mathbb{R}^3$ manifold for the basins of attraction of each equilibrium, and determines the ultimate outcome of the system given specific initial conditions. This is indeed the case: the line on the $x-$subspace such that $R_0(\bar{x}) = 1$ is the boundary at which the curvature of the schistosomiasis trajectory changes sign[9]. Figure 9 shows this phenomenon for two different magnitudes of $\alpha'(x)$: integrating numerically $\dot{m}_t$ in time for both m and x one can see that for the "higher" $\alpha'(x)$ there exists a threshold level of wealth $\bar{x}$ at which the R_0 falls below unity and the disease starts converging towards the disease-free equilibrium. Since $\alpha(x)$ is continuous and is at least $\mathcal{C}^1$, the subspace $\{\bar{x}, m, c\} \in \mathbb{R}^3$ is compact, and acts as separator for the two equilibria.

This is based on whether the initial condition is on which side of the separatrix surface $\{x, m, c\}$ that generates a $R_0(x) = 1$. This is because in an instant dt if for any initial state with optimal control $\mathbf{z}_0 = [x_0, m_0, c_0]$ we have $R_0(x_0) < 1$, the disease will tend to decrease towards zero, increasing instantaneous wealth and shifting the relative consumption level on the saddle path towards the disease-free equilibrium. Viceversa if the initial state generates a $R_0(x_0) > 1$. In the case of perfect equality on the R_0, the dominant eigenvalue will dictate in which basin of attraction will the system end up in, and has to be determined numerically. The separatrix surface is continuously dependent on the model parameters, and moves continuously in the $\{x, m, c\} \in \mathbb{R}^3$ space. If it crosses the disease-free equilibrium, implying that it generates a $R_0(x^*_{df}) > 1$, this causes one of its eigenvalues to switch sign due to the fact that $F_{xx}(x^*_{df}, 0)$ is small compared to the household's ratio of intertemporal utility substitution.

[9]The sign change happens assuming the initial state is below the endemic level: if the initial condition m_0 is above the endemic level, then the trajectory will decrease in any case for all levels of wealth, but asymptotically reaching zero if wealth surpasses the separator threshold.

The joint dynamics of schistosomiasis and wealth can therefore generate a **poverty trap**, depending on whether the initial conditions of $\{x_0, m_0\}$ and the associated c_0 put the household on a trajectory attracted to the endemic equilibrium. The escape boundary for such a trap is the surface $\{\bar{x}, m_t, c_t\}$ on the subspace of $\mathbb{R}^3$ where $R_0(x) = 1$, which intersects the disease-free equilibrium's two-dimensional stable manifolds. This trap is however not necessarily a consequence of the absence of an escaping path, such as in the case of the endemic equilibrium being a sink: if the farming household was aware of the joint dynamics of production and schistosomiasis, then if the trap is not entirely attracting (the endemic equilibrium being an attracting saddle) it could adjust consumption, for example by shifting part of it in the future in order to move the wealth towards the escape boundary and ultimately avoid being stuck in such an undesirable state. In other words, a possible trap can be generated by the farming households being *unaware* of the joint effects of development and disease diffusion. In a scenario of an "absolute" poverty trap, where the endemic state is attracting, the only ways to escape the disease-driven trap without changing the structural parameters would be to shock the initial state either by reducing substantially the disease intensity by large-scale eradication projects or increasing household wealth, to the level at which the initial reproduction rate $R_0(x_0)$ goes below unity. In the case of a "relative" trap, where the endemic state remains a saddle point, another strategy to avoid this state of poverty generated also by limited information would be to increase awareness of the population of the disease dynamics, such as the importance of clean water access and better sanitary conditions connected to higher household wealth.

6.2 The role of fluctuations: parametric noise

Let us now extend the evolution equation of schistosomiasis for the more realistic case in which some parameters vary in time, representing a natural source of system stochasticity due to the natural fluctuations in the environment. Let us for simplicity assume uncorrelated randomness in the parameters, and model this noise source as a Gaussian source. The evolution equation of schistosomiasis given by (11) now is defined as the nonlinear stochastic differential equation

$$\begin{aligned} dm_t &= \left(\alpha(x_t)\frac{NBHm_t}{BHm_t + \beta} - \gamma m_t\right)dt + \sigma m_t dW_t \\ &= (\alpha(x_t)g(m_t) - \gamma m_t)dt + \sigma m_t dW_t \end{aligned} \tag{24}$$

in the filtered probability space $(\Omega, \mathcal{F}, P)$, where W_t is the standard Brownian motion. The measure of schistosomiasis prevalence is now continuously perturbated by log-normal fluctuations. This is a natural consequence of the parametric noise: as an example, assume that at every instant t the death rate of schistosomes γ becomes $< \gamma + \sigma\xi_t >$, where ξ is white noise. The parametric fluctuations are therefore proportional to the level of schistosomiasis intensity m_t, and in differential form require an increment of the form $\sigma m_t dW_t$, which also allows intensity to always remain positive. Equation (24) has drift and diffusion coefficients uniformly Lipschitz continuous. The diffusion one is immediately verified, and for the drift we have

$$\lim_{x\to\infty} \alpha(x)g(m) = 0 \quad , \quad \lim_{x\to 0} \alpha(x)g(m) = \alpha g(m)$$
$$\lim_{m\to\infty} g'(m) = 0 \quad , \quad \lim_{m\to 0} g'(m) = \frac{BN}{\beta}$$

where $\alpha(x)$ is at least $\mathcal{C}^1$ by assumption. The last result, boundedness of the derivative, allows to say that for all $m, n \in \mathbb{R}^+$ there exists a $m_1 \geq g'^{-1}(BN/\beta)$ such that

$$\alpha(x)g(m) - \alpha(x)g(n) \leq \alpha(x)g'(m_1)(m-n)$$

for all x because of the mean value theorem. We then have

$$||g(m) - g(n)|| \leq K||m-n||,$$

where $K = g'(0)$, which is the condition of Lipschitz continuity for the drift, since for the $-\gamma m$ part it is trivially satisfied. This allows for a square-integrable, nonanticipating path m_t as the solution of (24) with initial condition m_0 for all values of the system $\mathbf{z} \in Z$, and this solution is unique.

The probability density of the disease intensity $P(m,t)$ is therefore continuously determined by the household's production and consumption decisions. The system defined by (24) and the joint household optimization problem, for an initial observation of $\mathbf{z}_0 = [x^0, m^0, c^0]$, is therefore fully characterized by the Kolmogorov forward equation with simultaneous dynamic constraints, and is a system of mixed ordinary and partial differential equations given by

$$\begin{aligned} \partial_t P(m,t) &= \partial_{mm}\frac{\sigma^2 m^2}{2}P(m,t) - \partial_m\Big[\alpha(x)g(m) - \gamma m\Big]P(m,t) \\ \partial_t x &= F(x,m) - \delta x - c \\ \partial_t c &= -\frac{u'}{u''}\left[F_x(x,m) - (\delta+\rho)\right]. \\ & \quad x_0 = x^0, c_0 = c^0, P(m,0) = \delta(m-m_0), \end{aligned} \tag{25}$$

where $\delta(.)$ is the Dirac mass function. This system is then solved forward in time. The non-stationary density has clearly no analytical solution, but writing (25) in non-divergence form sheds light on the probabilistic nature of the feedback effects of wealth on the disease distribution:

$$\begin{aligned} \partial_t P(m,t) &= \frac{\sigma^2 m^2}{2}\partial_{mm}P(m,t) - \Big[\alpha(x_t)g(m) - \gamma m\Big]\partial_m P(m,t) - \\ & \quad - \Big[\alpha(x_t)g'(m) - \gamma - \sigma^2 m\Big]P(m,t). \end{aligned}$$

This shows that the difference between the deterministic disease evolution rate $\partial\dot{m}/\partial m$ and the variance of the fluctuations determines the *probability flow* in the random dynamical system. To see this, remembering that the schistosomiasis intensity is exogenously observed by the household at every dt who immediately adjusts after each fluctuation in m, assume that the fluctuations start at the deterministic endemic equilibrium $\mathbf{z}_e^* = [x_e^*, m_e^*, c_e^*]$. In dt, probability starts evolving according to

$$\partial_t P(m,t) = \frac{\sigma^2 m^2}{2}\partial_{mm}P(m,t) - \Big[\alpha(dx_e^*)g(m) - \gamma m\Big]\partial_m P(m,t) - \Big[\gamma\Delta R_0(dx_e^*) - \sigma^2 m\Big]P(m,t), \tag{26}$$

as well as the dynamic constraints on wealth and consumption. We define dx_e^* as the instantaneous variation

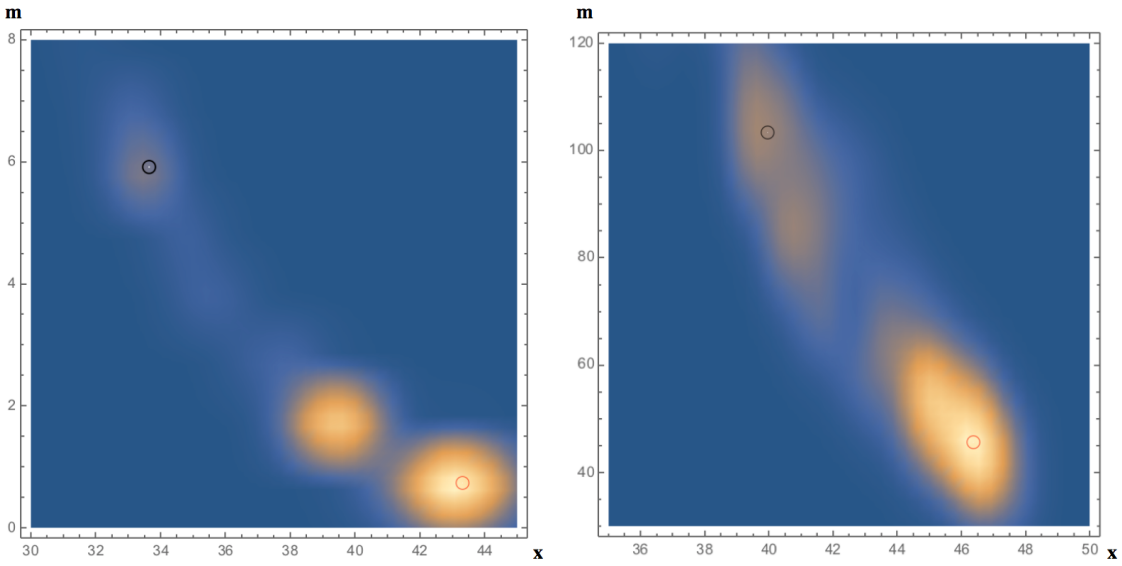

Figure 10: Monte Carlo simulated empirical density heatmap for joint wealth and disease states for initial conditions (black dots) not far from the deterministic steady states (red dots). We see evidence of the nature of the deterministic equilibria as pullback attractors. Left panel: disease-free equilibrium for $A = 1, \theta = 0.05$. Right panel: endemic equilibrium for $A = 1, \theta = 0.2$. For both we have $\sigma = 0.05$, elasticity of wealth 0.5, $\delta = 0.1$ and disease parameters chosen so that the natural R_0 is around 1.5. Production parameters used come from the empirical estimations. The empirical density is obtained by binning the data, limiting bin size to zero and estimating the joint density. 5000 simulations done with a Milstein scheme for the SDE and, for the two ODE, an exponential integration method for stiff systems.

of the "new" endemic equilibrium for wealth due to the fluctuations in m, since random dynamical systems are intrinsically non-autonomous. As m approaches zero[10] we have

$$\begin{aligned}\partial_t P(m,t) &= \frac{\sigma^2 m^2}{2}\partial_{mm}P(m,t) - \Big[\alpha(dx^*_{df})g(m) - \gamma m\Big]\partial_m P(m,t) \\ &\quad + \Big[\gamma(R_0(dx^*_{df}) - 1) + \sigma^2 m\Big]P(m,t).\end{aligned}$$

The flow of probability therefore is a result of the interplay between the magnitude of the fluctuations ($\sigma^2 m$) and the disease evolution rate, which asymptotically becomes the reproduction rate R_0. A $R_0(dx^*_e)$ greater than unity, net of fluctuation size, helps draw the probability towards a stationary measure centered on the deterministic endemic state $\mathbf{v}_e$ as initial condition for the household, while a $R_0(dx^*_{df})$ smaller than unity allows a household that starts near the disease-free state to be attracted towards a state of low disease intensity. This hints towards the nature of the two equilibria as *pullback attractors* (or random attractors), if the magnitude of the fluctuations is not too large, and the human-schistosome contact function as the key determinant of the stationary dynamics. Since if $m_0 \neq 0$ the boundary $m = 0$ is inaccessible, and so is $m = \infty$ because of the form of the drift on (24) and exploding trajectories of x_t are ruled out by the household's transversality conditions, we have that the stationary density for $\partial_t P = 0$, $P_s(m)$ is obtained by integrating twice the system (25) with a vanishing left-hand side, and is given by

[10]Lognormal fluctuations do not allow m to effectively hit the lower boundary.

$$P_s(m) = \mathcal{N}\frac{1}{\sigma^2 m^2}\exp\left(\int \frac{\alpha(F_x^{-1}(\rho+\delta)(\nu))g(\nu)-\gamma\nu}{\sigma^2\nu^2}d\nu\Big|_m\right) \tag{27}$$

where $\mathcal{N}$ normalizes the density to 1, and is a gamma-type distribution. We note that if the initial state of the disease is at the disease-free equilibrium, then $\mathbf{z}_e^* = [x*_e, 0, c_e^*]$ is in fact a steady state for both stochastic and deterministic system, since on that point $dm_t = 0$ for all t. The random dynamical system admits another attracting state with a stationary distribution, whose density is centered around the contact function evaluated at the deterministic long-run value of wealth. The deterministic equilibrium of household wealth and consumption, therefore, shapes the stationary disease density. The wealth process, being function of the disease intensity, also becomes a stochastic quantity, and because of the form of the production function (12) its stationary density is of the inverse gamma-type. Figure 10 presents Monte Carlo density estimates of the simulated system that shows that from initial conditions not too far from the deterministic steady states the stochastic system converges in probability to these states, if fluctuations are not too large. This is consistent with the nature of the equilibria as pullback attractors, with probability flow regulated by the (now stochastic) wealth-dependent R_0.

For the deterministic case, well-behavedness of the system and boundedness of its trajectories was guaranteed by the nature of the dynamics and the transversality condition. The stochastic case requires more attention, and let us therefore be more rigorous. The random dynamical system that defines the state $\mathbf{z}_t = [m_t, x_t, c_t]$ given by

$$\begin{aligned} d\ \mathbf{z}_t &= \mathbf{A}(\mathbf{z}_t)\times dt + \boldsymbol{\sigma}\times dW_t \\ \mathbf{A}(\mathbf{z}_t) &= [\dot{m}_t, \dot{x}_t, \dot{c}_t]' \\ \boldsymbol{\sigma} &= [\sigma m_t, 0, 0]', \end{aligned} \tag{28}$$

is defined on a complete metric space (Z, d), $Z \in \mathbb{R}^3$ equipped with an appropriate distance function d, with a σ-algebra over the filtered probability space $(\Omega, \mathcal{F}, P)$, and is a measurable map $\varphi : \mathbb{R}^+ \times \Omega \times Z \to Z$ given by $(t, \omega, z) \to \varphi(t, \omega)z$ such that

$$\begin{aligned} \varphi(0,\omega) &\quad \text{is an identity map, i.e. maps on itself} \\ \varphi(t,\omega) : Z \to Z &\quad \text{is continuous} \\ \varphi(t+s,\omega) &= \varphi(t,\theta_s\omega)\varphi(s,\omega)\ \forall t, s \in \mathbb{R}^+. \end{aligned}$$

The last property is known as the cocycle property, where θ_s is the Brownian shift, defined for each t as

$$\theta_t : \Omega \to \Omega, \omega \to \bar{\omega} \text{ s.t. } \theta_t\omega_s \to \bar{\omega}_s := \bar{\omega}_s = \omega(t+s) - \omega(t)$$

for all s in the sample space Ω. Attractors in this random system are therefore intrinsically connected with the nature of pullback attractors for non-autonomous systems[11], similarly to the "internal" non-autonomous nature of the household optimization problem, which considers the disease dynamics as exogenous time-

[11] To see this, interpret the presence of noise as a continuous exogenous modification of the deterministic system trajectories, thus effectively transforming the autonomous controlled system (15) to a non-autonomous one.

varying parameters. Following Crauel et al. (1997), a random set $\omega \to E(\omega)$ is said to be a random (pullback) attractor for (28) if it satisfies three conditions. First, it's a random compact set with respect to the probability measure P, implying that the distance on the metric space (Z, d) between any $z \in Z$ and $E(\omega)$ is measurable. This is easily seen to be satisfied since unbounded trajectories in x are excluded by the household's transversality condition and θ_t is by definition $\mathcal{F}$-measurable. The second is that E must be invariant, i.e. $\varphi(t, \omega)E(\omega) = E(\theta_t\omega)$ for all t. This property is confirmed by inspection of the structure of the noise source in (28), where the fluctuation of the Brownian motion is proportional to the system state: we have therefore $E(\theta_t\omega) = \mathbf{z}_{t+dt} = \dot{\mathbf{z}}_t + \mathbf{z}_t\theta_t = \mathbf{A}(\mathbf{z}) + z_t\theta_t = \varphi(t, \omega)E(\omega)$. The third property is the connection with pullback attraction, and requires that for every set $V \subset Z$ we have

$$\lim_{t\to\infty} d(\varphi(t, \theta_{-t}\omega)V, E(\omega)) = 0, \tag{29}$$

where we use the distance measure d appropriate for the relative metric space (Z, d)[12]. Time therefore is let go to minus infinity, hence the pullback nature of the attraction, and the distance between all past trajectories of the map defined by the system (28) and the attractor set E goes to zero. This property is what allows us to characterize the attracting set of the deterministic equilibria: it's a standard result that if the pullback attractor is bounded in the past then it can be characterised as the union of all global solutions that are bounded in the past (Carvalho et al., 2012). We can therefore transform the random system (28) in a deterministic non-autonomous system, by rewriting the equation of the disease evolution to

$$\dot{m}_t = \alpha(x_t)g(m_t) - \gamma_t m_t, \tag{30}$$

where $\gamma_t = \gamma - \sigma W_t$, together with the deterministic dynamics of wealth and consumption. We immediately notice that the disease-free equilibrium satisfies (29) trivially: for $m_0 = 0$ any fluctuation is set to zero, and limiting time to minus infinity continues yielding bounded trajectories identically equal to zero, and the system stays stably on the corresponding disease-free states x_e, c_e. For non-zero disease states, using Ito calculus we have to check the behavior for every t of the solution of

$$m_t = \lim_{s\to-\infty} m_s \exp\left(\int_s^t \alpha(x_r)\frac{g(m_r)}{m_r}dr - \left(\gamma - \frac{\sigma^2}{2}\right)(t-s) + \int_s^t \sigma dW_t\right).$$

with the simultaneous integrated equations of (28), which remain autonomous and we omit. The solution always exists as a measurable path, as shown before; the stochastic integral in this equation is an Itô integral and is simply equal to $W(t, \omega) - W(s, \omega)$ where W are realizations of standard Gaussian variables. Exploding wealth trajectories are excluded by the transversality condition, $g(.)$ has bounded derivative as shown above, m cannot hit the zero boundary via (24) because of the lognormal nature of the fluctuations and therefore the limit exists and is bounded. There therefore exists an unique pullback attracting global solution in $m > 0$, and the pullback attractor exists as the solution of (30) together with the household control system, only obtainable numerically. However, we can study the limits of the solution, which are given by

$$\begin{aligned}
\lim_{m\to 0}\int_0^t dm_t &= 0 \\
\lim_{m\to\infty}\int_0^t dm_t &= \exp\left(-\int_0^t \gamma_s ds\right)\left[m_0 + \int_0^t \exp\left(\int_0^s \gamma_r dr\right)\alpha(x_s)ds\right].
\end{aligned}$$

[12] Usually the Hausdorff semi-distance is used, defined as $d(A, B) = \sup_{a\in A}\inf_{b\in B}|a - b|$.

Since $m \to \infty$ (or more realistically when the disease intensity is much larger than the snail death rate β) implies $x \to 0$, we can write the last equation as $m_t = e^{-\int_0^t \gamma_s ds}\left(m_0 + \alpha N \int_0^t e^{\int_0^s \gamma_r dr} ds\right)$, where $\gamma_t = \gamma - \sigma W_t$. This pullback limit can be obtained explicitly:

$$\lim_{s\to-\infty} e^{-\int_s^t \gamma_r dr}\left(m_0 + \alpha N \int_s^t e^{\int_s^r \gamma_u du} dr\right) = \frac{\alpha N}{\gamma_t}.$$

This result points in the right direction but a last step is required: using Itô calculus[13] we can obtain the solution of the limiting SDE and then examine its backward limit, which is given by

$$\begin{aligned}\lim_{s\to-\infty} m_t &= \lim_{s\to-\infty} \exp\left(-\int_s^t \left(\gamma + \frac{1}{2}\sigma^2\right) dr + \int_s^t \sigma dW_r\right) \times \\ &\quad \left(m_s + \int_s^t \alpha N \exp\left(\int_s^r \left(\gamma + \frac{1}{2}\sigma^2\right) dr - \int_s^r \sigma dW_r\right) ds\right) \qquad (31) \\ &= \frac{\alpha N}{\gamma + \frac{1}{2}\sigma^2 - \sigma dW_t}.\end{aligned}$$

We note that this result implies that in order for the non-autonomous change of variable in (30) to yield the correct result, (24) has to be interpreted in the Stratonovich sense. The level $\alpha N/\gamma$ is easily seen to be the endemic state of the deterministic disease equation with large m when household wealth limits to zero. The random attractor for the limiting case is therefore the equivalent of the endemic deterministic one, subject to fluctuations and the Itô correction of the drift. This means that for all m the solutions of (30) exist, and for all t the random set $E(\omega) = [\mathbf{z}^*_{df}, \varphi(t,\omega)\mathbf{z}^*_e]$ defines the pullback attractors for the stochastic system (28). The zero-disease state is pullback-attracting and absorbing, since it is a non-random state unaffected by fluctuations. The endemic state is pullback-attracting and has a stationary density measure given by (27). If the state is at a disease-free state and is shocked to a nonzero level of disease, (26) shows that the probability flow between the states responds to the wealth-dependent (and thus stochastic) R_0 of the disease. This is because it effectively controls the drift of (24), and pushes the state towards its attracting level.

A rigorous analysis of non-stationary dynamics, as well as the potential emergence of stochastic resonance phenomena due to fluctuations, is a matter of great interest as well as a considerable challenge. Further expansions in this direction would be the inclusion of heterogeneity in the households, thus allowing (24) to have a drift dependent on the density of the wealth of different households, thus becoming a SDE of the McKean-Vlasov type. Another extension would be to include explicitly the stochastic dynamics in the optimization process and therefore have a mean-field game problem: we leave these potential developments to future research.

6.3 Disease dynamics as internal constraint: social optima

We now extend the model to one of social optima, where the household (or an external policymaker) wants to know the optimal policies necessary once the disease dynamics are explicitly accounted for. We furthermore include the effect of hired labor supply on the infection parameter α, now made dependent on aggregate household (village) wealth x_t and labor supply l_t, in order to represent human-water contact related to agricultural work, as well as being compatible with the evidence of disease contagion stemming from labor

[13] The proof is entirely standard and left to Appendix F.

migration. We model this mechanism with $\alpha(x_t, l_t)$, where $\alpha_x < 0$ and $\alpha_l > 0$, and the diagonal elements of its Hessian are $\alpha_{xx} \geq 0$ and $\alpha_{ll} \leq 0$. Intuitively, richer households as before experience a reduced possibility of infection, while an increase in labor supply on the farm increases it. Note that in the previous framework we could easily incorporate the same effect of labor supply on the human-schistosome contact function and obtain identical dynamics, as long as the function remains decreasing in x. We include the schistosomiasis dynamics in a model of household optimization, where consumption and labor supply need to be chosen optimally. The problem can be formalized as follows:

$$\begin{aligned} \max_{c,l} \quad & \int_0^\infty e^{-\rho t} u(c_t) dt \\ \text{s.t.} \quad & \dot{x}_t = F(x_t, l_t, m_t) - \delta x_t - c_t - w l_t \\ & \dot{m}_t = \alpha(x_t, l_t) g(m_t) - \gamma m_t. \\ & x_t, m_t \geq 0 \ \ \forall t \\ & [x_0, m_0] = [x^0, m^0] \in \mathbb{R}^+. \end{aligned} \tag{32}$$

where all parameters are the same as in the previous framework. If labor is supplied both internally and externally in an exogenous (i.e. non-optimized) proportion, one can simply substitute $wl^e l_t$, where l_e is the proportion of hired (external) labor.

Since $\alpha_{ll} \leq 0$, the Hamiltonian associated with this problem is concave in the controls and therefore the resulting c_t^*, l_t^* are optimal. The co-state variables λ_t and μ_t, respectively associated to the wealth and disease variables, can be obtained by a dynamic envelope approach, and are given by

$$\lambda_t = \frac{\partial}{\partial x} \int_t^\infty e^{-\rho s} u(c_s^*) ds \tag{33}$$

$$\mu_t = \frac{\partial}{\partial m} \int_t^\infty e^{-\rho s} u(c_s^*) ds \tag{34}$$

where c_t^* is the optimal path of consumption in feedback form, which is in turn function of the state variables. With standard methods one can show that the complete controlled system that describes the model dynamics can be written as:

$$\dot{x}_t = F(x_t, l_t, m_t) - \delta x_t - c_t - w l_t \tag{35}$$

$$\dot{m}_t = \alpha(x_t, l_t) g(m_t) - \gamma m_t \tag{36}$$

$$\dot{c}_t = -\frac{u'}{u''} \left[F_x - \frac{\alpha_x}{\alpha_l}(F_l - w) - (\delta + \rho) \right] \tag{37}$$

$$\begin{aligned} \dot{l}_t = & \frac{1}{(F_l - w)\frac{\alpha_{ll}}{\alpha_l} - F_{ll}} \left\{ \frac{u''}{u'}(F_l - w)\ \dot{c}_t + \left[F_{lm} - \frac{g'(m_t)}{g(m_t)}(F_l - w) \right] \dot{m} + \right. \\ & \left. + \left[F_{lx} - \frac{\alpha_{lx}}{\alpha_l}(F_l - w) \right] \dot{x}_t + \frac{\partial \dot{m}_t}{\partial l}(F_l - F_m - w) - (\gamma + \rho)(F_l - w), \right\} \end{aligned} \tag{38}$$

joined by the transversality conditions $\lim_{t\to\infty} \lambda_t x_t = 0, \lim_{t\to\infty} \mu_t m_t = 0$. For the sake of clarity we omit the arguments of F and α and their partial derivatives in (37) and (38). We first notice that a key quantity

in determining optimal labor and consumption dynamics is the deviation from the optimal separable level of labor, $(F_l - w)$. Evolution of optimal consumption, *ceteris paribus*, is unequivocally decreasing with an increase in θ (the parameter calibrating the effect of schistosomiasis) if labor is underutilized with respect to its separable level, i.e. $F_l \geq w$. This happens since $\alpha_x < 0, \alpha_l > 0$ and the gradient of F is decreasing in m. At the social optimum, therefore, we have a modified Keynes-Ramsey rule for consumption that includes the feedback dynamics of labor choice and wealth via the gradient of $\alpha(x, l)$, modulated by the deviation from a separable equilibrium. Equation (38) is more involved but has a straightforward interpretation: the optimally controlled labor choice is dependent on the evolution of both state variables as well as optimal consumption, and this co-movement is calibrated again by the deviation of labor from its separable level.

The schistosomiasis intensity as before has two equilibrium levels, a disease-free one where $m^*_{df} = 0$ and an endemic level given by $m^*_e(x_t, l_t)$, now also explicitly function of hired labor. The endemic level is an increasing function of labor and a decreasing function of wealth. We first notice that the disease-free equilibrium $m^*_{df} = 0$ implies $\partial \dot{m}/\partial l|_{m^*_{df}=0} = 0$ and an equilibrium level of wealth and labor such that $F_l = w$. This means that in absence of schistosomiasis, the problem becomes separable between input and consumption choices. This implies that the co-state variable μ_t related to the disease evolution is zero everywhere, and from (33) one can see that the optimal consumption level and its relative utility path is unaffected by the disease dynamics. Zero co-state implies that optimal labor is equivalent to the static problem, in which at every instant the household utilizes labor equal to its (disease-free) static level where the marginal productivity of labor is equal to the wage paid. At the disease-free equilibrium, therefore the problem collapses to a standard saddle path-stable consumption choice. As we can see from (37) we obtain the standard equilibrium derived from the Keynes-Ramsey rule for optimal consumption. The endemic steady state for labor as a function of wealth solves

$$\frac{\partial \dot{m}_t}{\partial l_t} = (\gamma + \rho) \frac{F_l(x_t, m^*_e(x_t, l_t), l_t) - w}{F_l(x_t, m^*_e(x_t, l_t), l_t) - F_m(x_t, m^*_e(x_t, l_t), l_t) - w}, \tag{39}$$

which since $F_m < 0$ shows how the socially optimal choice of labor deviates from its separable level by an amount proportional to the marginal decrease in agricultural production due to schistosomiasis. This result implies that the Nash-consistent outcomes generated by households optimizing without considering the externalities of production on the disease dynamics generate outcomes which are socially inefficient, and the deviation can be shown to be

$$\Delta l^*_{sep} = -\left(|F_m(x_t, m^*_e, l_t)| + (\gamma + \rho) \frac{\mu_t}{\lambda_t} \right). \tag{40}$$

Since in the Nash outcomes the disease co-state μ is by construction zero, we have that for social optima the optimal labor must include the externalities that labor itself has on the increased human-schistosome contact function. Using (33) and (34), we see that the deviation from separability in the household problem is proportional to the ratio of the marginal effects of wealth and disease on the future optimal consumption path.

An explicit analysis of the stability of the equilibria in the social optima case is difficult due to the complexity of the dynamical system given by (35)-(38). What we can obtain, however, is that for the disease-free equilibrium again the R_0 is a key quantity in determining the stability of the equilibrium, since it can be shown that the determinant of the Jacobian of the system linearized around the disease-free equilibrium, where $l^* = F_l^{-1}(w)$, is given by

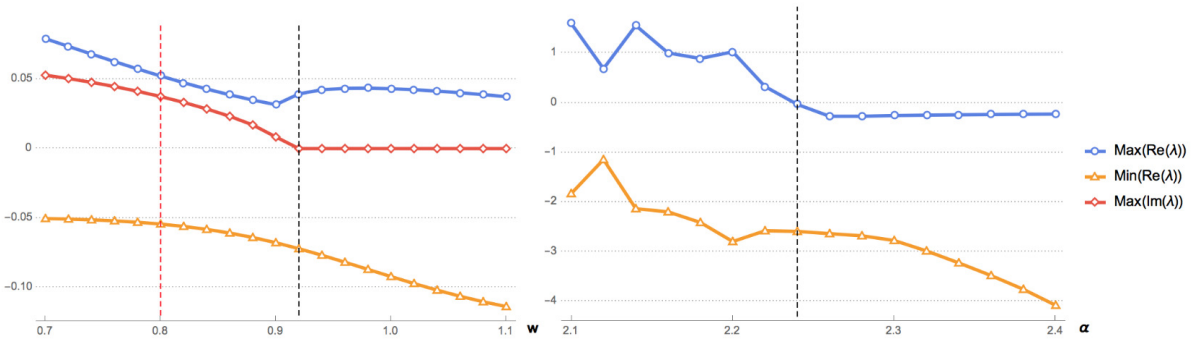

Figure 11: (Left panel) Bifurcation of the disease-free equilibrium due to an increase in wage, represented by the dashed black line. As the wage increases, semi-stable limit cycles vanish and the equilibrium becomes saddle-path stable. The dashed red line shows the level of wage that shifts the R_0 of the endemic equilibrium below the unity threshold, which is also where the dominant eigenvalue stops being a positive one. (Right panel) Emergence of an endemic state as a poverty trap, as the infection parameter α exceeds the threshold value indicated by the dashed black line. The imaginary part of the eigenvalues in this case is always zero and is therefore omitted from the figure.

$$\det J^*_{df} = \gamma(R_0(x^*_{df}, l^*_{df}) - 1)\left[-\frac{u'}{u''}\left(-F_{ll}(\gamma+\rho)\left(F_{xx} - \frac{\alpha_x}{\alpha_l}F_{lx}\right) + \gamma F_{xl}\left(F_{xl} - \frac{\alpha_x}{\alpha_l}F_{ll}\right)\right)\right]_{\mathbf{z}^*_{df}} \tag{41}$$

where $\mathbf{z}^*_{df} = [x^*_{df}, 0, c^*_{df}, l^*_{df}]$ is the equilibrium in the presence of no disease, and therefore the social optimum, and ΔR_0 is as before the percentage deviation of the disease reproduction rate from the threshold of unity. The exact nature of the (socially optimal) disease-free equilibrium requires eigenvalue analysis and cannot be explicited in closed form, although from (41) one can see how the sign of the determinant changes once the disease-free R_0 passes the threshold of unity. One can also see how bifurcations can emerge from parameter changes, since the argument of the determinant within square parentheses is a mixture of positive and negative quantities: however, specific conclusions require more specific assumptions on functional forms, although evidence with numerical simulations hints towards a greater instability of the system.

Bifurcations can therefore happen even at the disease-free equilibrium, something that was entirely ruled out in the Nash-outcome case, and the optimal equilibrium can exhibit partly stable limit cycles (i.e. limit cycles with both attracting and repelling manifolds), bifurcating from a standard saddle point. The endemic equilibrium shows evidence of bifurcations reciprocal to the disease-free ones: if one equilibrium switches from saddle to limit cycle, the other switches from limit cycle to saddle, and viceversa. It also shows evidence of absolute poverty traps, where all four eigenvalues are negative: however, the key difference with the Nash-outcome scenario is that whenever there exists an unstable manifold on the Pareto-dominated equilibrium, the social planner (or the representative household optimizing in a socially optimal way) can always choose labor and consumption accordingly in order to put the controlled state $\mathbf{z}$ on the unstable subspace and escape the trap. It is not possible unfortunately to flesh out the precise conditions for bifurcations and changes in the nature of the equilibria, due to the complexity of the system, but numerically it's a simple task to present some possible scenarios. For consistency we now assume the Cobb-Douglas production function given by (12) (total factor productivity $A = 1$, elasticities of wealth and labor equal respectively to 0.5

and 0.2, and $\theta = 0.1$ consistently with our empirical findings), logarithmic utility and contact function $\alpha(x_t, l_t) = \alpha(1 + l_t)/(1 + x_t)^s$, where $\alpha = 1, s = 0.5$. We further assume $B = 0.1, H = 0.1, N = 1, \beta = 1, \gamma = 0.2$, so that the "natural" R_0 of the disease (i.e. in absence of feedback effects) is above unity. The left panel of Figure 11 shows how as the exogenous wage increases, limit cycles around the socially optimal equilibrium disappear, and the fixed point becomes a standard saddle-path stable equilibrium which can now be reached with certainty. We notice that at the point on which the R_0 of the equilibrium changes sign, the dominant eigenvalue shifts from being positive to negative, implying that the net tendency of the system is to escape the disease-free state. In the socially optimal case, however, if the initial conditions do not force the farming household on the basin of attraction of the endemic equilibrium, the disease-free state can always be achieved by avoiding the unstable manifolds, adjusting consumption or labor (or both). In the right panel of Figure 11 we show the emergence of a poverty trap because of an infection parameter in the contact function α exceeding a certain threshold value, from where even a fully aware farming household cannot escape if the initial conditions are within its basin of attraction. One can notice that the dominant eigenvalue is always negative, therefore exhibiting a net tendency of the state to be attracted to the endemic equilibrium, but the existence of an unstable (i.e. escaping) manifold for an infection parameter not too large allows the household to choose either labor or consumption to escape. Once α passes the threshold indicated by the dashed black line in the figure, the trap is fully attracting. One also can notice that as the infection parameter increases, the more the Pareto-dominated state is asymptotically stable, as the dominant eigenvalue is negative and increases in absolute value. Needless to say, this analysis is entirely dependent on functional choices and therefore not as general as in the Nash-outcomes case, although it seems to be robust to alternative specifications.

7 Conclusions

In this paper I have studied the impact of schistosomiasis on agricultural yields in Burkina Faso by characterizing it as a productivity shock, and by creating a theoretical framework that is consistent with the empirical findings and allows to frame them in a wider perspective. By merging rich plot and household datasets with high-resolution disease maps, I have shown how schistosomiasis has a large, negative and nonlinear effect on agricultural productivity. Schistosomiasis therefore imposes a substantial burden on agricultural production, generating losses which range from an average of 6.6% to 32% for households and villages located in areas with the highest infection clusters. This burden is paradoxically reinforced both by poverty and by poverty-reducing measures such as water resources development. This mechanism is compatible with a poverty trap, and is explained by my theoretical findings where poorer households experience a higher disease reproduction rate (the R_0), leading to greater negative productivity shocks and increasingly lower levels of wealth and consumption. In the presence of environmental fluctuations, the endemic state becomes a global attractor of the resulting random dynamical system. In a scenario where farming households are not aware of the joint dynamics of agricultural production and schistosomiasis, we observe persistent states of poverty that in some cases could be avoided if complete information was available.

Given that grain output represents roughly 12% of Burkina Faso's gross domestic product, the preferred mean estimate implies that schistosomiasis is associated with economic losses which correspond to around 0.8% of GDP. From a policy perspective, perhaps the most interesting result is that while dams and reservoirs, *ceteris paribus*, increase agricultural yields, they can also induce substantial negative feedback effects by

spreading the disease. I show that this feedback is generated entirely by large-scale dams, which account for a significant portion of development finance. My work highlights how the study of the interactions between disease diffusion and economic development can benefit from the use of high-resolution data, which allows one to control for the numerous confounding factors that data at higher levels of aggregation necessarily miss. While the empirical focus of this paper has been on Burkina Faso, in part because it is likely to be a worst-case scenario due to the country's joint economic and epidemiological profile, this approach can be applied to any country in which schistosomiasis is endemic and, indeed, to the economic impact of many other diseases.

References

Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston (2020). Optimal targeted lockdowns in a multi-group sir model. *NBER Working Paper 27102*.

Acemoglu, D. and S. Johnson (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy 115*(6), 925–985.

Anderson, R. and G. Medley (1985). Community control of helminth infections of man by mass and selective chemotherapy. *Parasitology 90*(4), 629–660.

Antoci, A., M. Galeotti, and P. Russu (2011). Poverty trap and global indeterminacy in a growth model with open-access natural resources. *Journal of Economic Theory 146*(2), 569–591.

Audibert, M. (1986). Agricultural non-wage production and health status: A case study in a tropical environment. *Journal of Development Economics 24*(2), 275–291.

Audibert, M. (2010). Endemic diseases and agricultural productivity: Challenges and policy. *Journal of African Economies 19*(3), 110–165.

Audibert, M. and J.-F. Etard (1998, 07). Impact of Schistosomiasis on Rice Output and Farm Inputs in Mali. *Journal of African Economies 7*(2), 185–207.

Azariadis, C. and A. Drazen (1990). Threshold externalities in economic development. *The Quarterly Journal Of economics 105*(2), 501–526.

Azariadis, C. and J. Stachurski (2005). Poverty traps. *Handbook of Economic Growth 1*, 295–384.

Baldwin, R. E. and B. A. Weisbrod (1974). Disease and labor productivity. *Economic Development and Cultural Change 22*(3), 414–435.

Battle, K. E., T. C. Lucas, M. Nguyen, R. E. Howes, A. K. Nandi, K. A. Twohig, D. A. Pfeffer, E. Cameron, P. C. Rao, D. Casey, et al. (2019). Mapping the global endemicity and clinical burden of plasmodium vivax, 2000–17: a spatial and temporal modelling study. *The Lancet 394*(10195), 332–343.

Belloni, A. and V. Chernozhukov (2013, 05). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The Quarterly Journal of Economics 122*(1), 73–117.

Bleakley, H. (2010). Health, human capital, and development. *Annual Review of Economics 2*(1), 283–310.

Bleakley, H. and F. Lange (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics 91*(1), 52–65.

Bloom, D. E., D. Canning, and J. Sevilla (2004). The effect of health on economic growth: A production function approach. *World Development 32*(1), 1 – 13.

Boelee, E., P. Cecchi, and A. Koné (2010). *Health impacts of small reservoirs in Burkina Faso*, Volume 136. Iwmi.

Brooker, S., S. Whawell, N. B. Kabatereine, A. Fenwick, and R. M. Anderson (2004). Evaluating the epidemiological impact of national control programmes for helminths. *Trends in parasitology 20*(11), 537–545.

Carvalho, A., J. A. Langa, and J. Robinson (2012). *Attractors for infinite-dimensional non-autonomous dynamical systems*, Volume 182. Springer Science & Business Media.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Christiaensen, L., L. Demery, and J. Kuhl (2011). The (evolving) role of agriculture in poverty reduction—an empirical perspective. *Journal of Development Economics 96*(2), 239–254.

Crauel, H., A. Debussche, and F. Flandoli (1997). Random attractors. *Journal of Dynamics and Differential Equations 9*(2), 307–341.

De Janvry, A. and E. Sadoulet (2010). Agricultural growth and poverty reduction: Additiona l evidence. *The World Bank Research Observer 25*(1), 1–20.

Degarege, A., D. Degarege, E. Veledar, B. Erko, M. Nacher, C. M. Beck-Sague, and P. Madhivanan (2016, 12). Plasmodium falciparum infection status among children with schistosoma in sub-saharan africa: A systematic review and meta-analysis. *PLOS Neglected Tropical Diseases 10*(12), 1–18.

Diakité, N. R., M. S. Winkler, J. T. Coulibaly, N. Guindo-Coulibaly, J. Utzinger, and E. K. N'Goran (2017). Dynamics of freshwater snails and *Schistosoma* infection prevalence in schoolchildren during the construction and operation of a multipurpose dam in central Côte d'Ivoire. *Infectious Diseases of Poverty 6*(1), 93.

Didan, K. (2015). Mod13a2 modis/terra vegetation indices 16-day l3 global 1km sin grid v006 [data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-03-06 from https://doi.org/10.5067/MODIS/MOD13A2.006.

Ezeamama, A. E., A. L. Bustinduy, A. K. Nkwata, L. Martinez, N. Pabalan, M. J. Boivin, and C. H. King (2018). Cognitive deficits and educational loss in children with schistosome infection—a systematic review and meta-analysis. *PLoS Neglected Tropical Diseases 12*(1), e0005524.

Farmer, R. and J. Benhabib (1994). Indeterminacy and increasing returns. *Journal of Economic Theory 63*, 19–41.

Foster, R. et al. (1967). Schistosomiasis on an irrigated estate in east africa. iii. effects of asymptomatic infection on health and industrial efficiency. *Journal of Tropical Medicine and Hygiene 70*(8), 185–195.

Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data 2*(1), 1–21.

Goenka, A., L. Liu, and M.-H. Nguyen (2014). Infectious diseases and economic growth. *Journal of Mathematical Economics 50*, 34–53.

Gurarie, D. and E. Y. Seto (2009). Connectivity sustains disease transmission in environments with low potential for endemicity: modelling schistosomiasis with hydrologic and social connectivities. *Journal of the Royal Society Interface 6*(35), 495–508.

Hartgers, F. C. and M. Yazdanbakhsh (2006). Co-infection of helminths and malaria: modulation of the immune responses to malaria. *Parasite Immunology 28*(10), 497–506.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The elements of statistical learning*. Springer New York.

James, S. L., D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim, et al. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet 392*(10159), 1789–1858.

King, C. H. (2010). Parasites and poverty: The case of schistosomiasis. *Acta Tropica 113*(2), 95 – 104.

King, C. H. and M. Dangerfield-Cha (2008). The unacknowledged impact of chronic schistosomiasis. *Chronic Illness 4*(1), 65–79.

Lai, Y.-S., P. Biedermann, U. F. Ekpo, A. Garba, E. Mathieu, N. Midzi, P. Mwinzi, E. K. N'Goran, G. Raso, R. K. Assae, M. Sacko, N. Schur, I. Talla, L.-A. T. Tchuente, S. Toure, M. S. Winkler, J. Utzinger, and P. Vounatsou (2015). Spatial distribution of schistosomiasis and treatment needs in sub-saharan africa: a systematic review and geostatistical analysis. *The Lancet Infectious Diseases 15*(8), 927 – 940.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1), 67–80.

Matsuyama, K. (1991). Increasing returns, industrialization, and indeterminacy of equilibrium. *The Quarterly Journal of Economics 106*(2), 617–650.

Mbabazi, P. S., O. Andan, D. W. Fitzgerald, L. Chitsulo, D. Engels, and J. A. Downs (2011). Examining the relationship between urogenital schistosomiasis and hiv infection. *PLoS Neglected Tropical Diseases 5*(12).

McAdams, D. (2020). Nash sir: An economic-epidemiological model of strategic behavior during a viral epidemic. *Covid Economics (forthcoming)*.

Miguel, E. and M. Kremer (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Ouedraogo, H., F. Drabo, D. Zongo, M. Bagayan, I. Bamba, T. Pima, F. Yago-Wienne, E. Toubali, and Y. Zhang (2016). Schistosomiasis in school-age children in burkina faso after a decade of preventive chemotherapy. *Bulletin of the World Health Organization 94*(1), 37.

Perez-Saez, J., T. Mande, N. Ceperley, E. Bertuzzo, L. Mari, M. Gatto, and A. Rinaldo (2016). Hydrology and density feedbacks control the ecology of intermediate hosts of schistosomiasis across habitats in seasonal climates. *Proceedings of the National Academy of Sciences 113*(23), 6427–6432.

Perez-Saez, J., T. Mande, J. Larsen, N. Ceperley, and A. Rinaldo (2017). Classification and prediction of river network ephemerality and its relevance for waterborne disease epidemiology. *Advances in Water Resources 110*, 263–278.

Perez-Saez, J., T. Mande, and A. Rinaldo (2019). Space and time predictions of schistosomiasis snail host population dynamics across hydrologic regimes in Burkina Faso. *Geospatial Health 14*(2).

Poda, J., A. Traoré, and B. K. Sondo (2004). L'endémie bilharzienne au burkina faso. *Societé de Pathologie Exotique 97*(1), 47–52.

Richter, J. (2003). The impact of chemotherapy on morbidity due to schistosomiasis. *Acta Tropica 86*(2-3), 161–183.

Steinmann, P., J. Keiser, R. Bos, M. Tanner, and J. Utzinger (2006). Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases 6*(7), 411 – 425.

Takayama, A. (1985). *Mathematical Economics*. Cambridge University Press.

Udry, C. (1996). Gender, agricultural production, and the theory of the household. *Journal of Political Economy 104*(5), 1010–1046.

Walz, Y., M. Wegmann, S. Dech, P. Vounatsou, J.-N. Poda, E. K. N'Goran, J. Utzinger, and G. Raso (2015). Modeling and validation of environmental suitability for schistosomiasis transmission using remote sensing. *PLoS Neglected Tropical Diseases 9*(11), e0004217.

Wan, Z., S. Hook, and G. Hulley (2015). Mod11a1 modis/terra land surface temperature/emissivity daily l3 global 1km sin grid v006 [data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-03-06 from https://doi.org/10.5067/MODIS/MOD11A1.006.

Weisbrod, B. A., R. L. Andreano, R. E. Baldwin, E. H. Epstein, A. C. Kelley, and T. W. Helminiak (1974). Disease and economic development: The impact of parasitic diseases in st. lucia. *International Journal of Social Economics 1*(1), 111–117.

Woolhouse, M. (1991). On the application of mathematical models of schistosome transmission dynamics. i. natural transmission. *Acta Tropica 49*(4), 241–270.

World Bank Results (2017). Agriculture as a powerful instrument for poverty reduction.

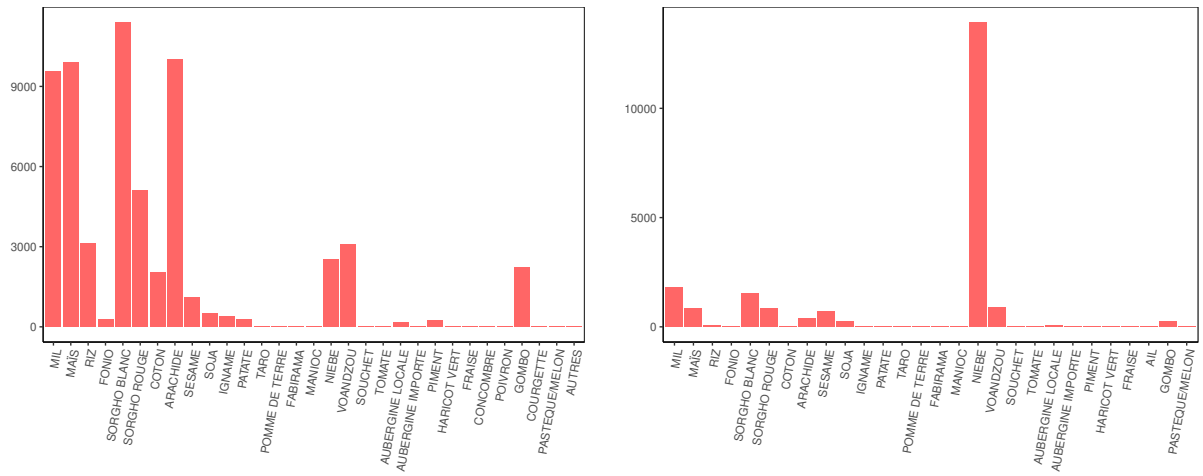

Figure 12: Crop choices for plot cultivations (above: main crop, lower: second crop, 2009 and 2011).

Appendix

A Data

A.1 Agricultural and Household data

The agricultural and household surveys we use has been obtained from the Direction Générale des Previsions et de Statistiques Agricoles at the Institut National de la Statistique et la Démographie in Ouagadougou (Burkina Faso), ranging 2003 to 2017. Overall the dataset contains 19,993 unique households located in 1,950 villages (see Fig.(1a) in the paper), cultivating 202,162 unique plots over the years. Each household on average cultivates around 10 plots throughout the years. Each plot can have up to two different crops, among 31 possible ones. The most common food crops are millet, corn, rice, sorghum and peanuts, and the only cash crop is cotton (traditional, bio, GMO). Figure 12 shows the counts of crop choices per yield in the 2009-2011 dataset. Most crops are food crops, and cash crops are almost entirely chosen as primary crop. As shown in Figure 5, the distribution of plot yield is identical after partialling out household fixed effects, implying heterogeneity is substantially generated at a plot level. Figure 13 shows how the villages cover substantially the area of Burkina Faso, and allow us to exploit a large variation in terms of both disease prevalence and geographical heterogeneity. Tables 1 and 2 show which households and villages are repeated across the 2010 cutoff, and are roughly equivalent in the years between 2009 and 2013. For the estimations of the paper we choose 2009 and 2011 for robustness, number of observed plots and model fit, although all results hold for other year combinations. Figure 13 presents the maps of the villages for the five relevant years around the 2010 cutoff (2009 to 2013) and their aggregated average yield.

A.2 Climatic, environmental and malaria covariates

Table 3 gives details on the data sources used in our analysis, and yearly means of the climate covariates are shown in Figure 14.

Table 1: Repeated households

	2011	2012	2013	2014	2015	2016	2017
2003	4	4	4	2	2	2	2
2005	5	6	6	3	3	3	3
2006	5	6	6	2	2	2	2
2007	5	6	6	3	3	3	3
2008	119	120	128	19	18	18	19
2009	3,040	2,994	3,185	10	10	10	11
2010	3,049	3,066	3,208	18	19	18	17

Table 2: Repeated villages

	2011	2012	2013	2014	2015	2016	2017
2003	18	18	19	16	16	16	16
2005	25	25	26	23	23	23	23
2006	25	25	26	21	21	21	21
2007	26	26	27	23	23	23	23
2008	651	648	697	235	233	235	236
2009	617	616	657	133	131	133	134
2010	588	586	618	125	124	125	126

Table 3: Details on remote sensing climatic and environmental covariates

	Website	Source	Covariates
Daily rainfall	CHIRPS-V2	Funk et al. (2015)	Yearly rainfall Dry-period rainfall Wet-period rainfall Mean length of dry spells Max length of dry spells
Monthly mean surface temperature	MODIS MOD11A1	Wan et al. (2015)	Mean yearly surface temperature (day) Mean yearly surface temperature (night)
Monthly mean air temperature		Perez-Saez et al. (2019)	Mean yearly air temperature Mean yearly night temperature Mean dry-period day temperature Days with temperature above seasonal median Days with temperature below seasonal median
Vegetation indices	MODIS MOD13A2	Didan (2015)	Mean yearly Normalized Vegetation Index (NDVI) Mean yearly Enhanced Vegetation Index (EVI)
Malaria	Malaria Atlas	Battle et al. (2019)	Prevalence of *Plasmodium Falciparum* (2003-2017)

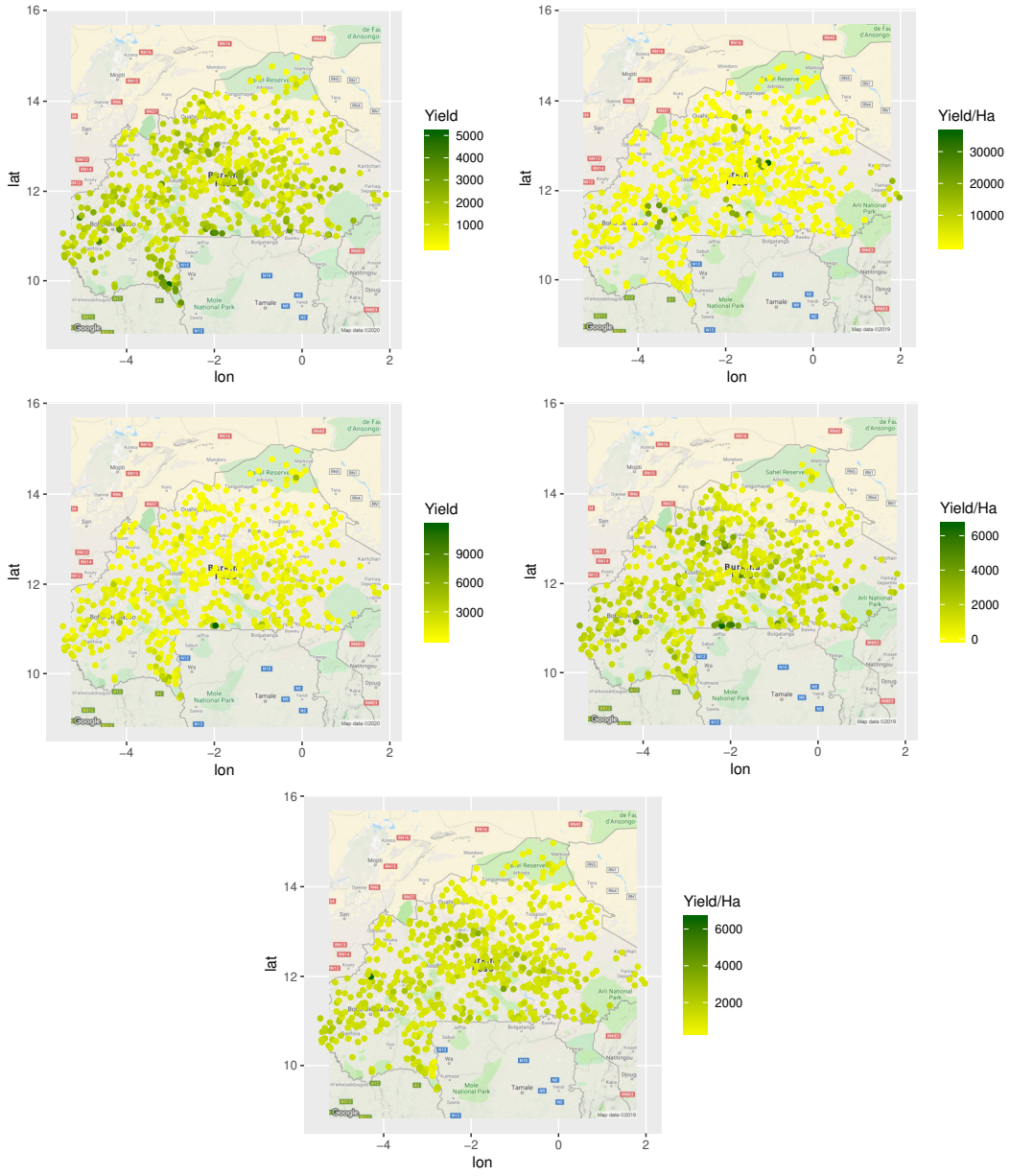

Figure 13: Geolocalized villages and relative average agricultural yield for the years 2009 to 2013 (from top left, in order)

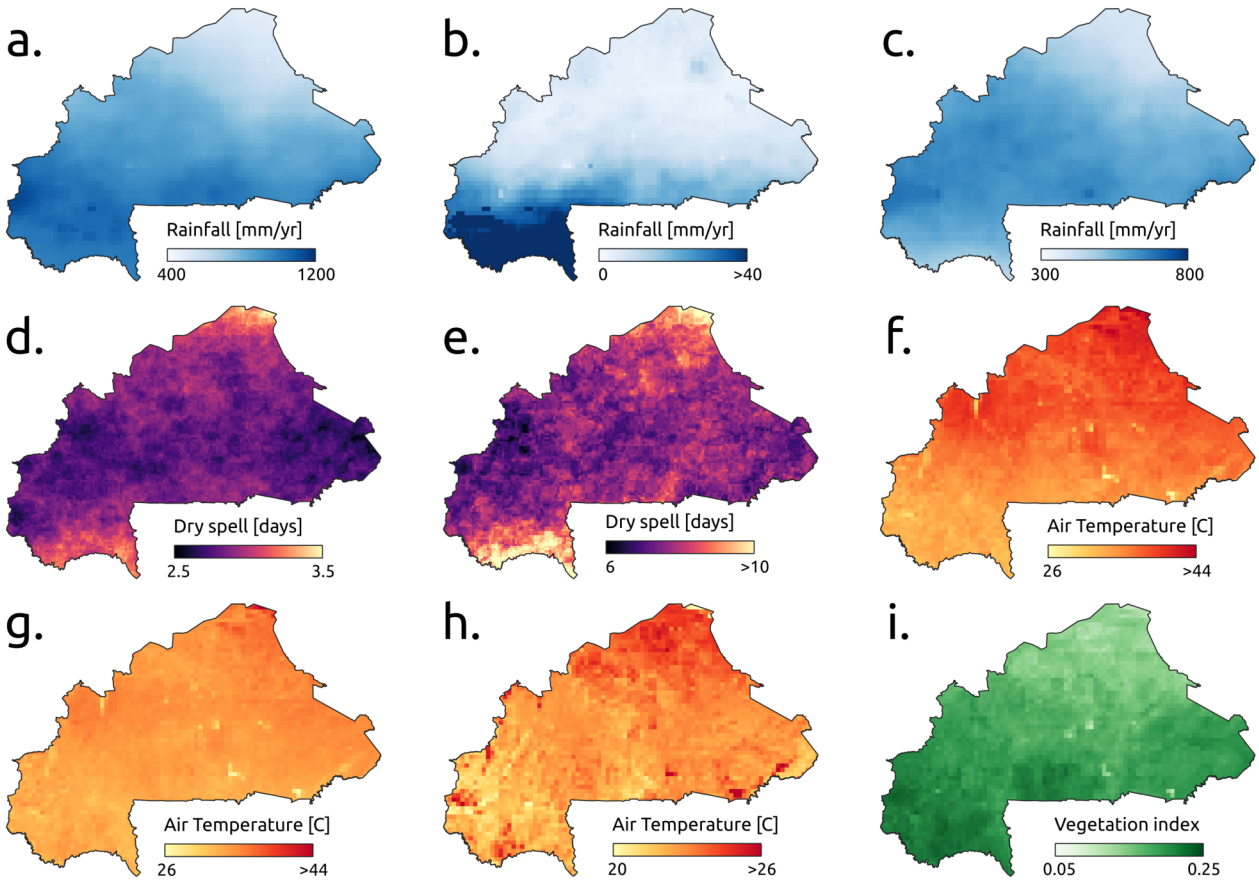

Figure 14: Climatic and environmental covariates used in the analysis. (a) Mean yearly precipitation, (b) mean dry-period precipitation (January to May), (c) Mean wet period precipitation (June to October), (d) Mean length of dry spells during the wet period. (e) Mean of maximum length of dry spells during the wet period, (f) Mean dry-period air temperature, (g) Mean day temperature, (h) Mean night temperature, (i) Mean Enhanced Vegetation Index (EVI). All figures show the yearly mean for the 2004-2018 period, while yearly rasters are included as covariates.

B Schistosomiasis and agriculture

B.1 Transforming schistosomiasis prevalence into infection intensity

The input to our analysis consists of modelled estimates of schistosomiasis prevalence, obtained via Bayesian geostatistical methods, over a pixel grid at 5×5 km. The first map is obtained directly from Lai et al. (2015), and the second is estimated in the same way. However, disease burden is more directly related to infection intensity (worms per person), rather than prevalence. We therefore proceed to transforming modelled prevalence into infection intensity using common assumptions about the distribution of worm burden in the human population.

The data consists of prevalence and mean intensity of infection from parasitological surveys in Burkina Faso coming from the Schistosomiasis control program of the Ministry of Health, of which part of the data is published in Ouedraogo et al. (2016). All samples correspond to school-aged children. We here make the assumption that the number of *Schistosoma* eggs per sample in the population follows a negative binomial distribution Anderson and Medley (1985), Brooker et al. (2004). We have that the prevalence is given by: $p = 1 - P(X = 0) = 1 - \left(1 + \frac{\mu}{k}\right)^{-k}$. Parasitological data across endemic countries suggest that the aggregation parameter k is not constant across transmission settings, but rather varies as a function of the mean intensity of infection μ in the population (mean eggs/person) Brooker et al. (2004). Following other studies, we test for either a constant, linear ($k(\mu) = a + b\mu$) or a quadratic ($k(\mu) = a + b\mu + c\mu^2$) relation. Inference on $k(\mu)$ is drawn through maximum likelihood estimation. For a given parameters set θ, the likelihood of the parameters the parasitological data is given by a binomial distribution on the number of infected people $n_{infected,i}$ among the sampled population $n_{sampled,i}$ in each village i among m sampled villages:

$$\mathcal{L}(\theta, \mathcal{D}) = P(\mathcal{D}|\theta) = \prod_{i}^{N} p_i^{n_{infected,i}} (1 - p_i)^{n_{sampled,i} - n_{infected,i}},$$

where p_i is the prevalence of schistosomiasis in village i. The negative log-likelihood is then minimized using standard optimization algorithms. We select for the best model using Akaike Information Criterion (AIC) $\text{AIC} = -2\log(\mathcal{L}) + 2m$, where m is the number of parameters in the model. We find the that there is a strong support for a non-constant functional form (Table 4). The best-fitting quadratic parameter is not statistically significant, so the linear form is retained.

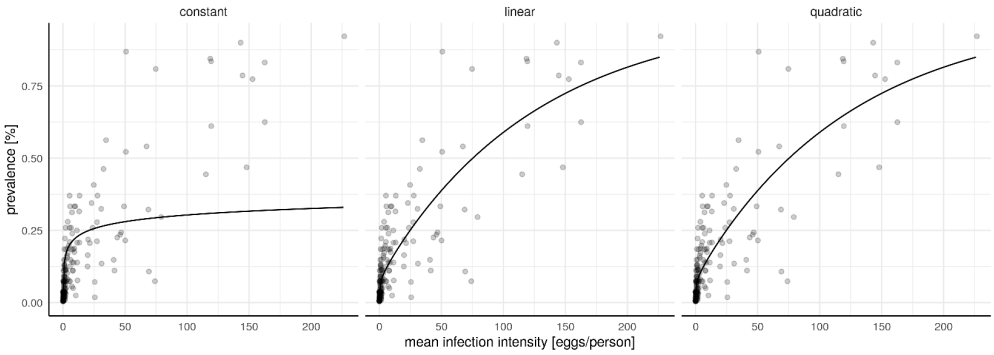

Figure 15: Functional forms of the variation of parasite aggregation with mean infection intensity in Burkina Faso. The best fitting relationship is a linear function of mean egg-intensity (Table 4).

Table 4: Model selection results for the functional form of the aggregation parameter for the inversion of schistosomiasis prevalence into infection intensity.

$k(\mu)$	$\log(\mathcal{L})$	AIC	parameters		
			a	b	c
constant	-6682.02	13366	0.047	-	-
linear	-6020.48	12043	0.016	0.0012	-
quadratic	-6020.48	12043	0.016	0.0012	1e-12

B.2 Schistosomiasis and optimal input allocation

We now present a simple model of optimal household input allocation with a distortion to labor input due to schistosomiasis. We posit the disease to be a shock to *effective* labor supply E, of the form

$$E = (\phi^{-\tilde{\theta}} L),$$

with $\theta > 0$, and that the total agricultural output of the household is generated by

$$Y = A(\phi^{-\tilde{\theta}} L)^{\alpha} F(X, K),$$

where A is total factor productivity, $\alpha < 1$ and $F(.)$ is the remainder of the production technology, dependent on k inputs$X$ chosen by the household (pesticides, livestock use) and inputs K which affect production but cannot be chosen (climate, land type). A household aims to maximize its instantaneous agricultural profits by choosing optimally$L$ and X, and is given by

$$\arg\max_{L,X} \ \Pi \ = \ p^y A(\phi^{-\tilde{\theta}} L)^{\alpha} F(X, K) - wL - \sum_{i=1}^{k} p_i X_i \tag{42}$$

where p_i are the input prices, w is the labor wage, and p^y is the price of output. The $k+1$ optimal allocations are then obtained by solving the system of $k+1$ first-order conditions:

$$L^*(A, \phi^{-\tilde{\theta}}, K, w, p), X_i^*(A, \phi^{-\tilde{\theta}}, K, w, p) \rightarrow \begin{cases} L^* = \left(\frac{\alpha p^y A F(X^*, K)}{w \phi^{\alpha\tilde{\theta}}}\right)^{\frac{1}{1-\alpha}} \\ X_i^* = F^{-1}\left(\frac{p_i \phi^{\alpha\tilde{\theta}}}{A L^*}\right) \end{cases}$$

where $p = (p^y, p_1, \ldots, p_k)$ is the vector of prices. Assuming an interior maximum, depending on the properties of $F(.)$, the log-linearized optimal production is therefore given by

$$y^* = \tilde{A} - \theta\phi + \alpha L^*(A, \phi^{-\tilde{\theta}}, K, w, p) + \tilde{F}(X_i^*(A, \phi^{-\tilde{\theta}}, K, w, p), K),$$

which recovers the quasilinear structure of the equation (2) of the paper. If F is assumed Cobb-Douglas, then the model becomes fully linear.

The total marginal effect of schistosomasis on log-linearized optimal production, therefore, is given by

$$\frac{dy^*}{d\phi} = -\theta + \alpha\frac{dL^*}{d\phi} + \sum_{i=1}^{k}\frac{\partial F(X^*,K)}{\partial X_i^*}\frac{dX_i^*}{d\phi},$$

where $\theta = \alpha\tilde{\theta}$, and the specification becomes identical to Eq. (1) (i.e. schistosomiasis is a pure productivity shock) if $\frac{dL^*}{d\phi}$ and all the $\frac{dX_i^*}{d\phi}$ are zero. Note that this model can straightforwardly be extended to a more general form $F(L, X, K)$, but we choose this specification to better illustrate the choice of estimating Eq. (1) and choosing both log-linear and adaptive forms for F.

The condition of zero derivatives is what we observe in the data: a first set of reduced-form evidence comes from studying the interactions of schistosomiasis on all input variables. We obtain a clear sign of zero signal: Table 6 reports the coefficients of the interactions, and omits both the rest of the controls as well as the interaction of schistosomiasis intensity on non-chosen inputs (climate, land type). Table 5 shows how for all of the choice input variables the marginal impact of schistosomiasis is not significant, hence validating our initial assumption of identifying schistosomiasis as a productivity shock in estimating Eq.(1). The results are robust to more flexible specifications (interactions, adaptive methods), as well as by instrumenting the intensity with the snail densities. All other controls used for each line of5 are omitted. Unlike schistosomiasis, malaria does not seem to have a significant effect in this specification as a productivity shock, which we find to be a believable finding: malaria, by the nature of its health effects and burdens, is more consistent with a temporary shock to direct labor supply. This identification would have to be tested with a completely different strategy, which we leave to future research.

B.3 Full results and robustness checks

We present here the full set of results and robustness checks for the estimation of the overall impact of schistosomiasis on agricultural yield. As stated in the main paper, we specify the production

$$Y_{ihjt} = A_{hj}(1+\phi_{hjt})^{\theta}F(X_{ihjt})e_{ihjt},$$

where Y_{ihjt} is the yield (output per hectare) of plot i, farmed by household h in village j at time t, ϕ_{hjt}, which is common to all plots cultivated by a given household, represents the direct effect of schistosomiasis on productivity, calibrated by a parameter θ. The term $(1+\phi)$ is proxied with the measure of infection intensity in terms of mean egg-output per person discussed in B.1, denoted by $I_{jt} = (1 + Int_{jt})$. Taking logarithms then results in the quasilinear model (2), where our goal is to identify the parameter θ. In all estimations, for the variables that include zeroes (particularly plot surface and yield) we use the inverse hyperbolic sine transformation instead of the natural logarithm. We begin by imposing a Cobb-Douglas functional form on $F(.)$, which yields a fully linear model. Details on the machine learning architecture are the following: in general we use a 5-fold cross-validation for every method, and the best method is chosen by means of a mean squared error criterion, in the non-IV case is gradient boosting machines and for the IV case random forests. For random forests, we use 1000 trees, each with a minimum terminal node size of 5. Tuning is done by randomly sampling input variables for each split, with a step factor of 1.5 with improvement of 0.05, starting from the square root of the number of sampled columns. Gradient boosting machines are fit using a Gaussian distribution since in all cases both y and I are continuous, with 1000 trees each with a shrinkage parameter of 0.1, a within-fold cross-validation of 3 folds and 20% of the observations in the training set chosen to fit each subsequent tree. Neural networks work less well than the other two methods,

Table 5: Input use and schistosomiasis intensity

y	*Intensity (Clustered s.e.)*
log plot surface	0.0003 (0.0002)
presence of hired labor	0.0003(0.0004)
presence of motorized labor	0.0008(0.0008)
presence of manual labor	0.0002(0.001)
presence of horse-driven labor	-0.001 (0.0007)
presence of labor from mutual aid	-0.002(0.001)
total agricultural livestock	0.0001(0.001)
total livestock	0.0001(0.0008)
cows	-0.0011(0.0021)
horses	-0.0024(0.0039)
pigs	-0.0014(0.0028)
goats	0.001(0.0021)
self-consumption	0.0018(0.0014)
cows sold	0.0002(0.0015)
cows bought	-0.0042(0.0028)
pigs sold	0.0004(0.002)
pigs bought	0.0008(7e-04)
goats sold	0.0002(5e-04)
goats bought	-0.0001(0.0012)
gifts	0.0001(0.001)
thefts	0.0011(0.001)
type of plot farming	-3e-04(0.0012)
npk (kg)	0.0023(0.0016)
urea (kg)	-0.0025(0.0015)
phosphates (kg)	0.0006(0.0006)
solid pesticides (kg)	0.0015 (0.001)
liquid pesticides (cl)	0.0014(0.0012)
herbicides (g)	0.0025(0.0036)
herbicides (cl)	-0.0014(0.0032)
fungicides (g)	-0.0017(0.0023)
fungicides (cl)	0.0008(0.0006)
rodenticides (g)	0.0036(0.0025)
rodenticides (cl)	0(0.0013)
number of household members	-0.011(0.0042)
number of families in the household	0.0006(0.0007)
number of young members working	0.0045(0.0026)
number of women working	0.0052 (0.0029)
rainfed agriculture	0.0001(0.0001)
Observations	51,055
Time, crop and household f.e., village clustering	*p<0.1; **p<0.05; ***p<0.01

Table 6: No signal on schistosomiasis interactions for choice inputs

	y: Log Yield
intensity×log surface	−0.001 (0.003)
intensity×presence of hired labor	−0.005 (0.003)
intensity×presence of motorized labor	0.008(0.008)
intensity:×presence of manual labor	−0.012(0.011)
intensity×presence of labor from mutual aid	0.002 (0.003)
intensity×total agricultural livestock	0.001(0.001)
intensity×total livestock	0.00000(0.00004)
intensity×cows	0.0003 (0.0003)
intensity×horses	0.007(0.004)
intensity×pigs	−0.0001(0.0004)
intensity×goats	0.0001(0.0003)
intensity×self-consumptions	0.0001(0.0003)
intensity×cows sold	−0.001 (0.001)
intensity×cows bought	−0.001(0.001)
intensity×pigs sold	−0.00001(0.001)
intensity×pigs bought	0.001 (0.001)
intensity×goats sold	−0.0001(0.001)
intensity×goats bought	0.00005 (0.001)
intensity×gifts	−0.001(0.0005)
intensity×thefts	0.0001(0.0001)
intensity×type of plot farming	0.003 (0.002)
intensity×npk (kg)	0.00000(0.00003)
intensity×urea (kg)	−0.0001(0.0001)
intensity×phosphates (kg)	−0.0001(0.0002)
intensity×solid pesticides (kg)	−0.00001(0.00001)
intensity×liquid pesticides (cl)	0.00000(0.00000)
intensity×herbicide (g)	0.00000 (0.00000)
intensity×herbicide (cl)	0.00001 (0.00001)
intensity×fungicide (g)	−0.00004(0.00005)
intensity×fungicide (cl)	−0.00001 (0.00001)
intensity×rodenticide (g)	0.00000(0.00001)
intensity×rodenticide (cl)	0.00002(0.00002)
intensity×number of household members	0.0001(0.001)
intensity×average age	−0.0001 (0.0002)
intensity×number of families in the household	−0.0001(0.006)
intensity×number of children	0.001(0.001)
intensity×number of young working	−0.0003 (0.001)
intensity×number of women working	−0.0003(0.001)
Observations	51,055
R^2	0.336
Adjusted R^2	0.290
Residual Std. Error	1.209 (df = 48072)
Time, crop and household f.e., village clustering	*p<0.1; **p<0.05; ***p<0.01

although the estimates are not statistically different, but result in a greater MSE. They are fitted with 5 units per hidden layer with maximum allowable weights of 5000 with 1% decay and 1000 maximum iterations.

Aggregating at the village level, the quasilinear model for a representative household is of the form:

$$y_{jt} = \tilde{A}_j - \theta I_{jt} + \tilde{F}(X_{jt}) + \alpha_t + \epsilon_{jt}. \tag{43}$$

All village-level regressions are obtained assuming a Cobb-Douglas specification for $F(.)$, since village aggregation yields reduced-form estimates. The model is then estimated with standard panel IV techniques in the same manner as was the case at the plot-level. When controlling for spatial effects and spatial correlation, the model we fit after stacking observations and imposing a Cobb-Douglas functional form is the mixed spatial autoregressive model (SAR) given by:

$$y_{jt} = (1_{2j} - \rho W_j)^{-1}[I_{jt}\theta + \tilde{X}_{jt}\beta + \alpha_t + \epsilon_{jt}], \tag{44}$$

where W_j is the matrix of spatial weights given by the distance between the j villages, ρ is the spatial autoregressive parameter, and $\tilde{X}_{jt}$ is the matrix of factor inputs expressed in logarithmic form. We fit a model where both dependent and independent variables are spatially lagged. In order to check for the presence of spatial random effects, we fit a model with a random term $u(d)$ with mean zero and spatial covariance function $M_\nu(d)$. We adopt the standard choice of a Matérn covariance function, given by

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu \left(\sqrt{2\nu}\frac{d}{\rho}\right).$$

where d is between-village distance, ν and ρ are positive parameters, K_v is the modified Bessel function of the second kind and Γ is the Gamma function. All parameters are then fit by maximum likelihood, allowing us to estimate the spatial correlation between village pairs, and to obtain its rate of decay as distance increases.

We present results for households and villages observed in 2009 and 2011. These years are chosen to maximize the number of observations and provide the best fit and accompanying model diagnostics. Results are broadly similar for other combinations of years for which there are enough repeated observations.

Figure 16 reports the different estimates of θ. Given that the intensity measure is expressed in terms of worm eggs per person, and that its value is in the 0-110 range, the coefficient associated with intensity is expected to be negative and of the order of 10^{-2}-10^{-3}. The point estimate represents the marginal impact of one additional worm egg per person on log yield.

The upper three estimates in Figure 16 impose a Cobb-Douglas functional form, yielding a fully linear model; the first includes year and crop fixed effects, the second adds household ones. The point estimates imply a loss in mean yield of between around 1% and 4%, with villages in the top 5% quantile of disease intensity losing between 5% and 20%; however, the coefficients are not statistically significant at usual levels of confidence. While time- and crop-invariant unobservables are controlled for, endogeneity and measurement error remain a concern. Since our measure of disease prevalence/intensity is model-based, the associated prediction error is included as a covariate, thereby hopefully reducing measurement error associated with the key RHS variable. However, since our intensity measure is constant at the village level, and because individual households may display heterogeneous levels of infection within a same village, measurement er-

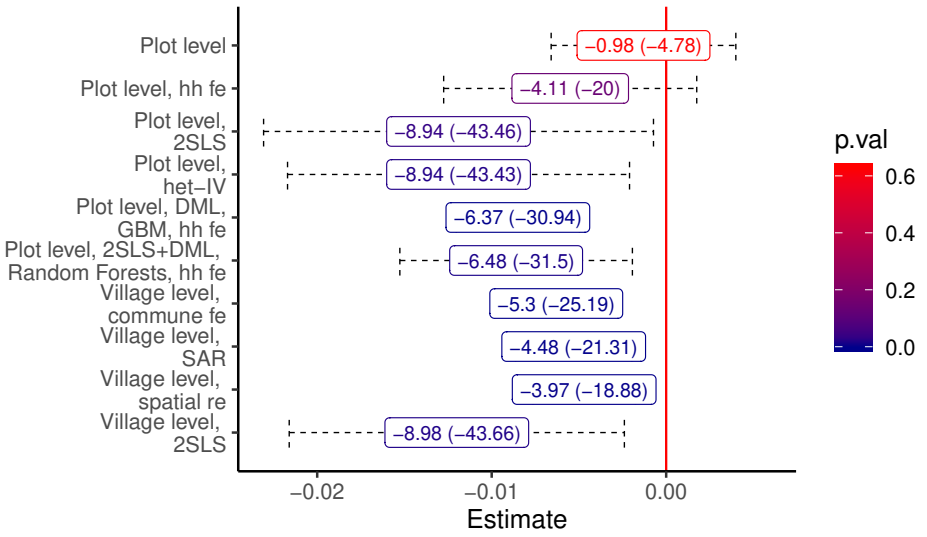

Figure 16: Estimates of the loss in yield due to schistosomiasis for 2009 and 2011 (95% confidence interval). Each label reports average and top 5% percentile losses. All estimations include the full set of controls and time fixed effects, and plot-level estimations also include crop fixed effects. Errors are clustered at village level (commune level for commune fe), and cluster-bootstrapped for 2SLS.

ror in all likelihood remains. As such, we then proceed to instrument disease intensity with the density of the freshwater snails that serve as intermediate hosts of schistosomes, using densities of both snail species *Biomphalaria* and *Bulinus* in different seasons (dry, rainy and winter) in order to instrument both forms of the disease as described in the paper. The third entry in Figure 16 shows how instrumenting intensity results in the point estimate of mean yield loss rising to 8.48% (40.99% in the top 5% intensity quantile). Estimation done via a control function approach yields identical estimates and also allows to test whether schistosomiasis intensity is actually endogenous, and this hypothesis is not rejected. It is then of interest to inquire what could the source of this endogeneity be, whether from the likely presence of measurement error or from a wider set of causes including omitted variable bias. We can get a clearer idea of the cause by using heteroscedasticity as an internal instrument for the mismeasured value of schistosomiasis intensity in the log-linear (Cobb-Douglas) model, a two-step GMM procedure first presented by Lewbel (2012). The necessary heteroscedastic covariance restriction arises both by means of the nested structure of the data and by natural occurrence. The results are almost identical to the 2SLS estimates, as shown in the fourth line of Figure 16, and thus offers a possible explanation on the sources of the endogeneity bias in the standard panel regressions. Reliance on internal instruments, however, whilst controlling for coefficient biases, does not necessarily warrant causality in the effect, and therefore we rely on the instrumented specifications (log-linear and machine learning) as the main results of the paper.

Once we relax the functional form assumption on the production technology and adopt DML techniques, point estimates of the marginal effect of schistosomiasis intensity fall slightly, but precision improves. Indeed, as is apparent in Figure 16, all DML 95% confidence intervals are entirely contained within the confidence interval of the corresponding quasi-linear model, once household fixed effects are accounted for. For our 2009-2011 sample, our preferred non-instrumented model is given by gradient boosting machines, and results in a mean yield loss of 6.22% (30.05% in the top 5% intensity quantile). All three adaptive methods

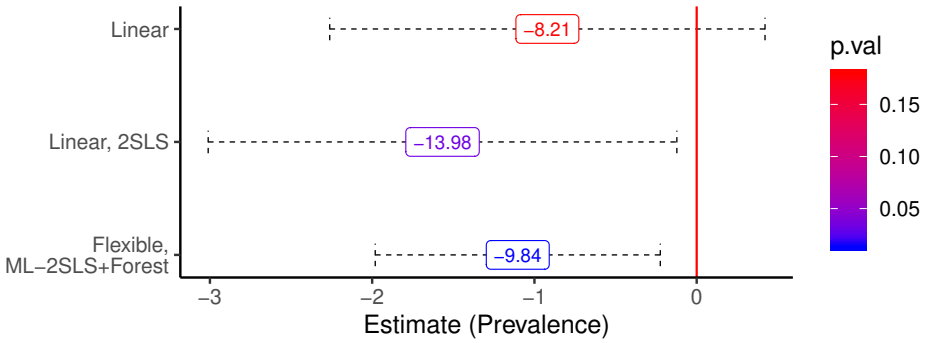

Figure 17: Estimates of the impact of schistosomiasis prevalence

(random forests, gradient boosting machines and deep neural networks) perform similarly, and differences in MSE between the three are marginal. The IV DML estimates, which are computationally expensive in that they require at least 2-fold cross validation, yield slightly larger marginal effects, which are again estimated quite precisely. The instrumented DML estimate implies a mean loss of yield of 6.63% (32.04% in the top 5% intensity quantile). Figure 17 shows the robustness of the results to using directly the measure of schistosomiasis prevalence instead of intensity: the standard panel regression remain statistically insignificant, and the instrumented ones are not statistically different from the intensity ones but higher in terms of estimated percentage loss of yield (9.5% instead of 6.48%). The estimates reported in the main paper are therefore on the conservative side: the reasons behind this slight variation could lie in the direct dependence of the prevalence measure from the climate controls used in the geostatistical model estimation.

As expected, village-level estimates are more precise: within-cluster correlation is dampened and measurement error is reduced. Estimation of the linear model with commune fixed effects yields a loss in yield of 5.38% (25.57% in the top 5% intensity quantile), while the IV results yield a loss of 9.94% (47.26% in the top 5% intensity quantile). Aggregating at the village level also allows us to study the impact of spatial correlation, with the key issue being whether controlling for higher (commune) level unobservables is sufficient to control for spatial correlation. We begin by estimating Moran's I for the models with only time fixed effects as opposed to time and commune fixed effects: spatial correlation falls from 0.13 - low, but statistically significant at all conventional levels of confidence ($p = 2e^{-5}$) to a statistically insignificant -0.07 ($p = 0.970$). Commune level unobservables therefore seem to lie at the root of any spatial correlation present in the data. Fitting a spatial lag model with a weight matrix obtained from distances between villages yields a similar point estimate of 4.48%.[14] The map in the right-hand panel of Figure 5 provides a graphical illustration of the weight matrix. We then check for the presence of spatial random effects. Point estimates are slightly lower, with an average yield loss of 4%, but the null that this is equal to previous estimates cannot be rejected. We then construct a measure of spatial correlation from the parameters obtained from the spatial estimations. The right-hand panel of Figure 5 shows that spatial correlation decays rapidly: it vanishes almost completely at a distance of 1 degree (110 km at the equator), and by 0.33 degrees (ca. 37 km) it already falls below 0.1. This confirms the previous result that commune-level fixed effects already remove most spatial dependence in the data. For all instrumented estimations we use both a Hausman-Wu test and a control function approach. Table 7 shows the relevance of the instruments (the first stage), while Table 8 reports the significance of the residuals of the first stage in the structural equation, which implies

[14]The choice of queen or rook-type adjacency weights does not significantly affect the estimates.

rejection of the hypothesis of exogeneity.

Table 7: First Stage

	Dependent variable: Schistosomiasis Intensity
	intensity
Biomphalaria, rainy season	2.063***
	(0.205)
Biomphalaria, dry season	0.970**
	(0.423)
Bulinus	0.204***
	(0.077)
Observations	51,055
R^2	0.422
Adjusted R^2	0.420
Residual Std. Error	9.439 (df = 50894)
Note: Other covariates omitted. Errors clustered at village level	*p<0.1; **p<0.05; ***p<0.01

Table 8: Rejection of exogeneity. Control function approach

	Dependent variable: Log Yield
First Stage Residuals	0.012**
	(0.006)
Observations	51,055
R^2	0.180
Adjusted R^2	0.178
Residual Std. Error	1.302 (df = 50895)
Note: Other covariates omitted. Errors clustered at village level	*p<0.1; **p<0.05; ***p<0.01

B.4 Non-linearity and interactions

As a robustness check, we fit a spline for the log-linear estimation in different combination of years, and Figure 18 shows how the results hold. There is no capping out effect as in the figure in the paper, because the households in high-intensity areas (100+ worm eggs/person) that drive that effect are either trimmed or not surveyed in the above year combinations. We also fit a spline using all years (2003-2017) and the results stay consistent, although likely plagued with a variety of issues, since there are effectively only two data points per village in time. The lower left panel of Fig.18 shows the good residuals fit for the log-linear estimation of Figure 2 in the paper. There is a small *positive* effect of schistosomiasis on agricultural yield for low levels of intensity, with the significantly negative marginal effect beginning at an infection intensity of 30 eggs/person. This small positive marginal effect for low intensity stems from the interaction

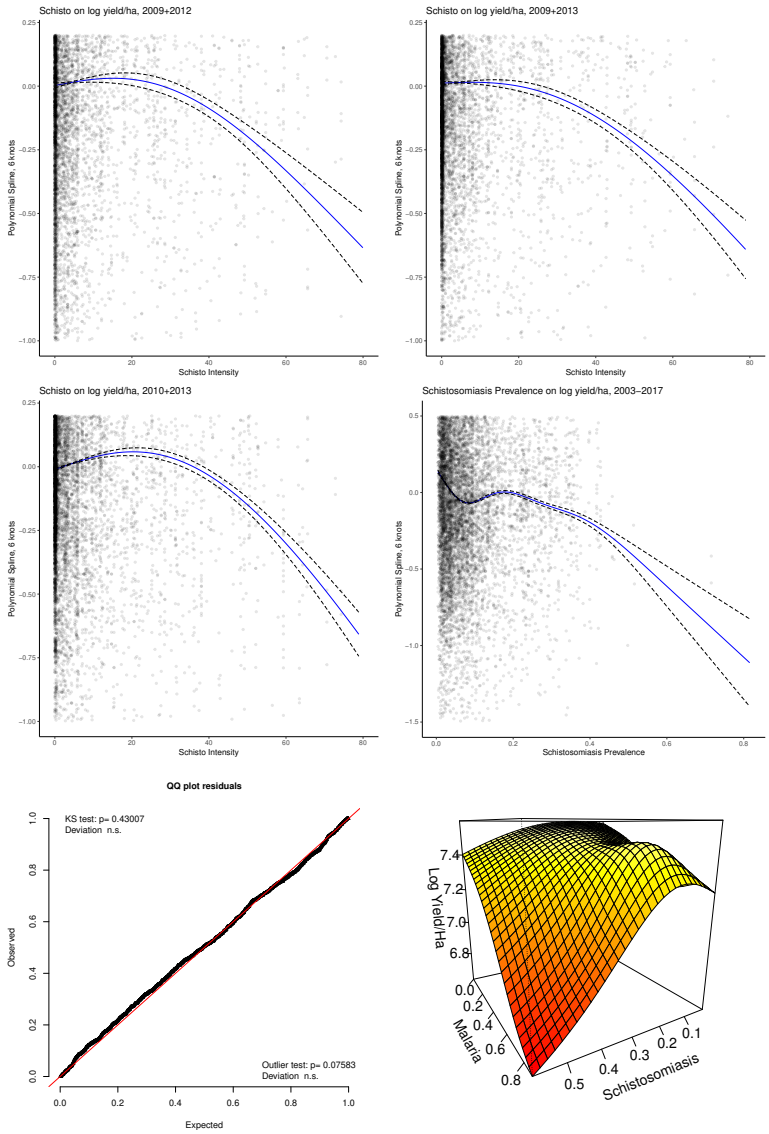

Figure 18: (Top) Spline fit for 2009, 2012 and 2013. (Middle) Spline fit for 2010, 2013, and using all years. (Bottom) Residuals fit for the 2009-2011 years (village level), and interaction of schistosomiasis and malaria prevalence measures.

between schistosomiasis and malaria: given that malaria is only available as a prevalence measure, we fit a multidimensional spline for the interaction between malaria and schistosomiasis *prevalence*. The relationship is highly nonlinear, as displayed in the right hand panel of Figure 18: high prevalence levels for both diseases correlate with low crop yields, but low to intermediate levels of schistosomiasis mitigate the negative effects of high malaria prevalence. The immunological mechanisms behind this phenomenon have been discussed by Hartgers and Yazdanbakhsh (2006) and Degarege et al. (2016), who present evidence on how low- to mid-intensity schistosomiasis can protect from acute malaria infections.

B.5 Other determinants of agriculture in the instrumented linear regressions

Table 9 reports the rest of the coefficients for the linear estimations. For clarity we only report the ones significant at standard levels of confidence: because of the nature of the data, constant at a village level, and the procedure being 2SLS, errors are cluster-bootstrapped at a village level, and therefore increase largely estimation errors and many controls result statistically insignificant. All continuous covariates are in logs. Signs are usually as expected, with an interesting negative quadratic relationship between plot surface and yield, suggesting decreasing returns to scale in plot size. Figure 19 shows nonlinear estimates of the correlation of schistosomiasis and the size of the cultivated plot. The negative quadratic relationship between yield and plot size remains established, and it seems that middle-sized plots are the ones that generate the higher yield.

C Schistosomiasis and poverty: subsistence farming and crop choices

We investigate whether schistosomiasis affects households differently based on whether they cultivate cash crops (cotton) or food crops (all the rest, mostly of subsistence). The question is of substantial policy relevance: if schistosomiasis exerts a lower burden for cash crops, then policies aimed at poverty reduction and diversification of the agricultural sector will carry along a beneficial effect on the reduction of the economic impact of schistosomiasis. We first note that plots that cultivate cash crops are substantially larger than the rest: Figure 20 shows the density of the log surface of plots used for cotton and the overall surface density, and shows how plots that cultivate cotton as a cash crop are substantially larger than the others. Larger farms in Burkina Faso involve richer households, especially when agriculture is not anymore of the subsistence variety. Cotton, furthermore, is drought-resistant and mostly rainfall-fed, therefore less in need of plots being around large water reservoirs or networks. We expect the effect to be reduced for these crops: we also expect the productivity shock to be less harmful for farms and households where constraints are less binding and agricultural yield is not as directly crucial for survival. The upper panel of Figure 21 estimate the main plot-level model with the same techniques as the main estimation. Errors are cluster-boostrapped. Instrumentation of the model increases the estimates by a large amount, because subsetting to cotton crops reduces greatly the variation in reservoir area of the snail hosts, as well as their estimated density: IV estimates are then large and extremely noisy. In any case, there does not seem to be a significant adverse effect of schistosomiasis on households farming cotton as cash crop, and the reason is likely to be twofold. The first reason is in cotton being a drought-resistant crop and thus requiring less irrigation networks: in the paper we show how this directly reduces the effective burden of the disease. The second reason lies in the fact that farms and households farming cash crops own substantially larger plots and are on average much richer; this can have multiple implications, from increased access to clean running water to better sanitary conditions. As the lower panel of Figure 21 shows, food crops, and particularly smaller plots

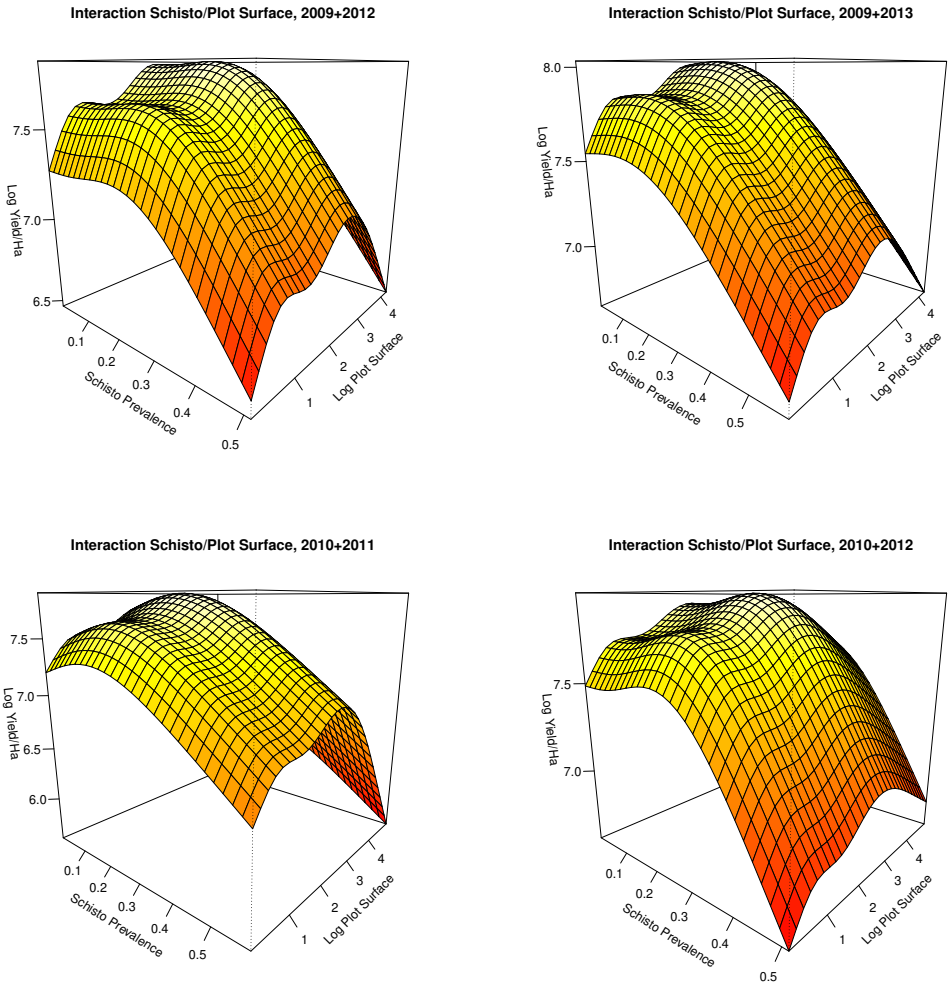

Figure 19: (Upper panel) Interaction of schistosomiasis intensity and plot surface (Other panels) Yield and plot surface for different levels of disease intensity

Table 9: Other determinants of agricultural yield

	Dependent variable: log yield
Log surface	0.466^{***}
	(0.054)
Log surface^2	-0.140^{***}
	(0.025)
Presence of hired labor	0.106^{***}
	(0.033)
Total livestock used in agriculture	0.057^{***}
	(0.022)
Female livestock used for breeding	-0.051^{**}
	(0.024)
Cows	0.029^{*}
	(0.015)
Sales of cows	-0.049^{*}
	(0.026)
Npk (kg)	0.020^{**}
	(0.008)
Urea (kg)	0.028^{***}
	(0.009)
Solid pesticides (g)	0.027^{*}
	(0.014)
Herbicides (g)	0.015^{*}
	(0.008)
Plot near houses	0.604^{**}
	(0.283)
Plot near bushes	0.577^{**}
	(0.283)
Plot near encampment	0.587^{**}
	(0.287)
Plot on flatland	0.185^{***}
	(0.040)
Plot farmed with half-moon technique	-0.665^{***}
	(0.203)
Dead/live hedges	0.169^{**}
	(0.080)
Maximum level of education in household: Non-alphabetized	-0.123^{**}
	(0.057)
Maximum level of education in household: Alphabetized	-0.057^{*}
	(0.034)
Maximum level of education in household: Primary	-0.162^{**}
	(0.073)
Loss due to flooding	-0.982^{***}
	(0.122)
Loss due to fire	-0.549^{***}
	(0.117)
Loss due to drought	-0.680^{***}
	(0.044)
Loss after harvesting	-0.987^{***}
	(0.156)
Mean temperature	-0.04^{*}
	(2.597)
Temperature in dry season	0.03^{**}
	(1.586)
Precipitation in dry season	0.721^{***}
	(0.227)
Max dry spell	-0.307^{***}
	(0.113)
Mean of land surface temperature (day)	-0.02^{**}
	(1.088)
Mean of land surface temperature (night)	0.022^{***}
	(0.915)
Enhanced Vegetation Index (level)	4.8^{**}
	(2.014)

Note: Time+hh+crop fe. Cluster-bootstrapped (village) *p<0.1; **p<0.05; ***p<0.01

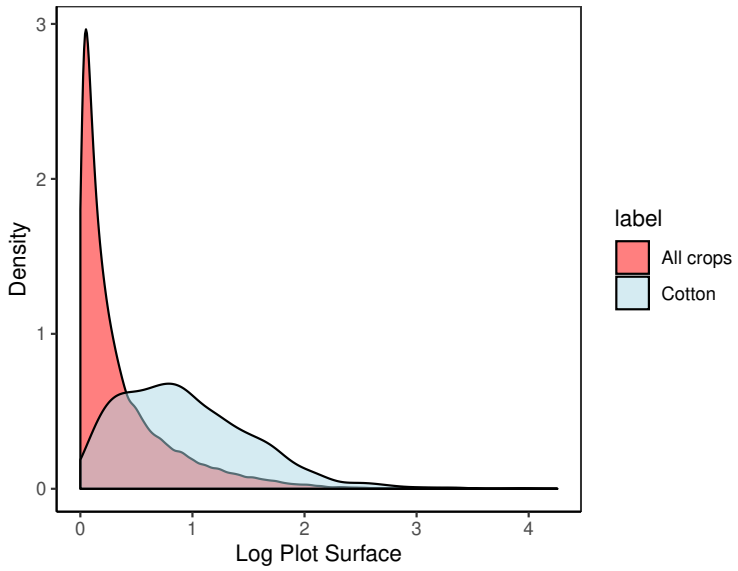

Figure 20: Distribution of plot sizes for cotton crops vs. all crops

mostly dedicated to farming of the subsistence variety, seem to suffer most of the disease burden, and hint at schistosomiasis being indeed a disease of poverty. Economic development aimed at raising individuals and households from having to farm for survival to a standard of agriculture with better returns and conditions will generate both an improvement in living standards and a reduction of the disease burden.

D Schistosomiasis and poverty: burden heterogeneity

We now turn our attention to the heterogeneity of the disease burden for different joint quantiles of plot surface and plot yield. In the previous section we showed how cotton crops - cash crops - are on average much larger, and how for these crops the burden of schistosomiasis is small and not statistically significant. Cash crops are usually farmed by households which do not rely on subsistence agriculture, own larger plots and are thus on average richer. As stated in the paper, this is a first sign of how schistosomiasis causes and reinforces poverty, but it might be that smaller, poorer households also use cash crops (particularly GMO-resistent varieties of cotton) in order to reduce the risk of adverse shocks. This is not a common occurrence in our dataset, as seen in the histograms of Fig. 12. Table 10 shows the coefficients of the interaction term of the specifications stated in the main paper, and Table 11 reports the respective p-values. These coefficients represent the *added* effect of schistosomiasis for the plots that are at the tail of the joint surface-weight distribution, for different quantiles. We note that the effect stops being significant around the 55th percentile. We then focus on households farming only food crops that lie at the very bottom of the joint distribution *between households*, grouping total harvest weight and overall plots surface per household and creating an indicator for the households at the 5th and 10th bottom percentiles. These households exhibit the lowest overall production and farm smallest plots, and thus belong to a situation of extreme poverty, especially given the overall state of agriculture in 2009-2011 Burkina Faso. Table 12 shows the coefficients for the respective quantiles, the diagonal of which is represented in the paper. All estimates are from a log-linear model where schistosomiasis intensity is instrumented with a control function approach (standard 2SLS yields equivalent results), and errors are cluster-bootstrapped at a village level.

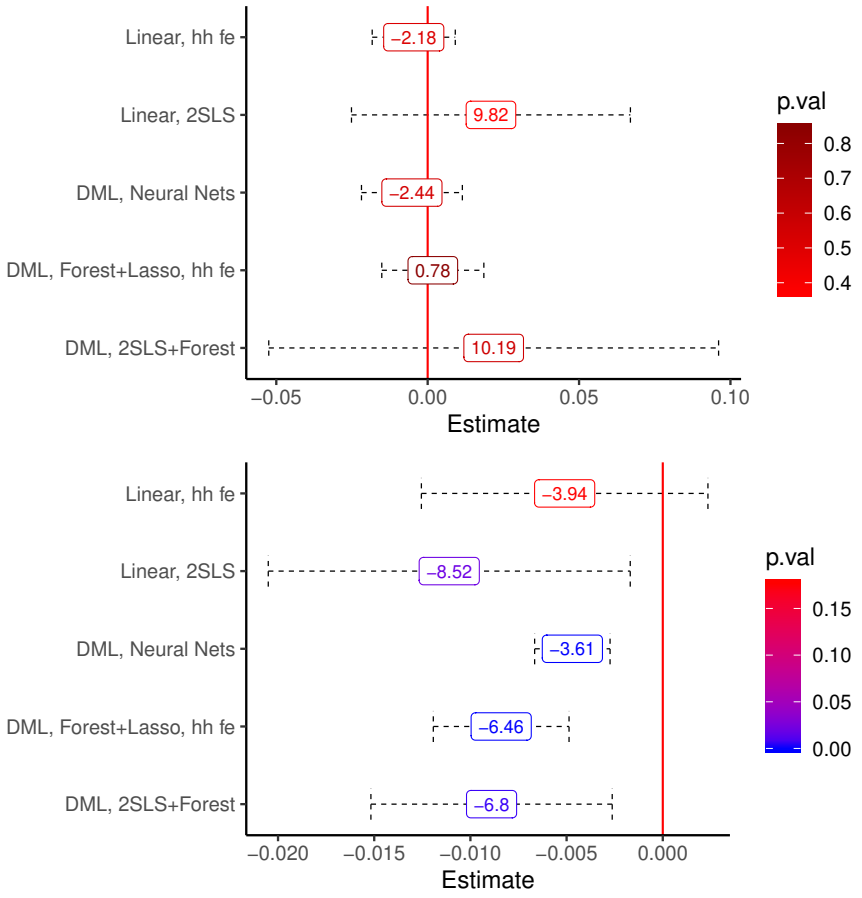

Figure 21: (Upper panel) Schistosomiasis burden for cash crops-only plots: no effect (Lower panel) Burden for food crops, mostly of subsistence: large effect. In labels: percentage yield loss for villages at the average schistosomiasis intensity.

Table 10: Estimates of the interaction term for the poverty quantile and schistosomiasis intensity. Rows: surface, Columns: yield

	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
0.2	-0.015	-0.013	-0.011	-0.012	-0.013	-0.015	-0.015	-0.015	-0.014	-0.015	-0.014
0.25	-0.015	-0.013	-0.011	-0.012	-0.013	-0.014	-0.014	-0.013	-0.013	-0.013	-0.012
0.3	-0.013	-0.012	-0.011	-0.011	-0.011	-0.012	-0.011	-0.011	-0.010	-0.010	-0.010
0.35	-0.013	-0.012	-0.011	-0.010	-0.011	-0.011	-0.011	-0.010	-0.010	-0.010	-0.009
0.4	-0.012	-0.010	-0.009	-0.009	-0.009	-0.009	-0.009	-0.008	-0.008	-0.008	-0.007
0.45	-0.011	-0.010	-0.009	-0.008	-0.008	-0.009	-0.008	-0.007	-0.007	-0.007	-0.007
0.5	-0.010	-0.009	-0.008	-0.007	-0.007	-0.008	-0.007	-0.007	-0.006	-0.007	-0.006
0.55	-0.009	-0.008	-0.007	-0.006	-0.006	-0.007	-0.006	-0.006	-0.005	-0.006	-0.005
0.6	-0.007	-0.007	-0.005	-0.005	-0.005	-0.006	-0.005	-0.005	-0.005	-0.005	-0.005
0.65	-0.006	-0.006	-0.004	-0.004	-0.004	-0.005	-0.004	-0.004	-0.004	-0.004	-0.004
0.7	-0.005	-0.004	-0.003	-0.003	-0.003	-0.004	-0.003	-0.003	-0.003	-0.003	-0.003

Table 11: P-values for the interaction term for the poverty quantile and schistosomiasis intensity. Rows: surface, Columns: yield

	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7
0.2	0.012	0.014	0.021	0.011	0.004	0.002	0.002	0.002	0.003	0.001	0.001
0.25	0.008	0.011	0.016	0.009	0.004	0.002	0.003	0.003	0.005	0.002	0.002
0.3	0.011	0.010	0.012	0.013	0.006	0.004	0.005	0.005	0.010	0.005	0.005
0.35	0.010	0.010	0.009	0.012	0.006	0.004	0.005	0.005	0.009	0.005	0.005
0.4	0.019	0.019	0.019	0.018	0.014	0.010	0.014	0.016	0.023	0.017	0.017
0.45	0.025	0.022	0.024	0.024	0.021	0.015	0.019	0.022	0.031	0.023	0.021
0.5	0.028	0.026	0.026	0.025	0.022	0.016	0.021	0.024	0.031	0.025	0.025
0.55	0.045	0.042	0.049	0.047	0.038	0.029	0.036	0.040	0.051	0.042	0.041
0.6	0.081	0.070	0.091	0.080	0.062	0.047	0.059	0.058	0.067	0.061	0.055
0.65	0.119	0.097	0.162	0.138	0.125	0.101	0.132	0.118	0.131	0.107	0.105
0.7	0.174	0.183	0.264	0.236	0.211	0.160	0.185	0.157	0.182	0.148	0.150

Table 12: Coefficients of the interaction term for household at the bottom quantiles of total yield and plot surface (***: significant at 1%.).

	lower 5% surface	lower 10% surface
lower 5% yield	-0.044***	-0.039***
lower 10% yield	-0.049***	-0.045***

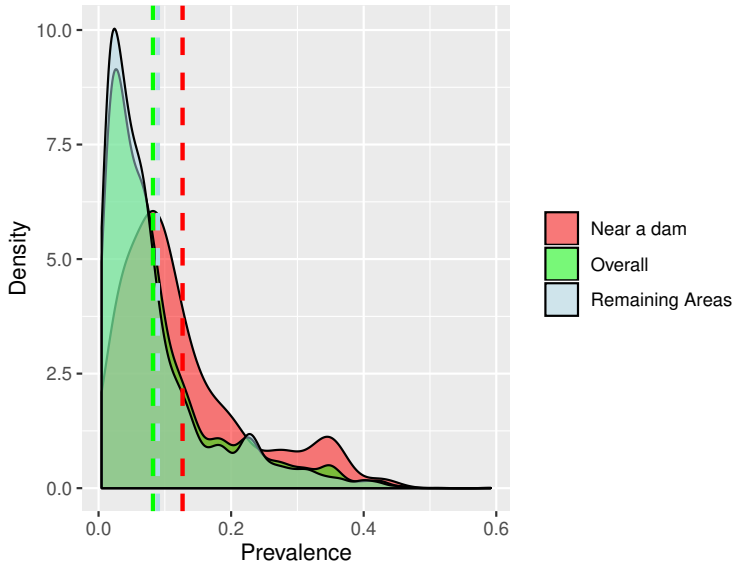

Figure 22: Distribution of schistosomiasis prevalence in areas near a large dam, plotted against the overall distribution and the distribution of the disease in the rest of the country.

E Schistosomiasis and water resources development

We report here the full set of results for the analysis of schistosomiasis and water resources development. We focus on Burkina Faso's four main dams:the Bagré Dam in Boulgou province, which also involves Zoundwéogo, Kouritenga and Ganzourgou provinces, the Kompienga Dam in Kompienga province, the Ziga Dam in Oubritenga province and the Léry Dam in Nayala province. First we notice how the distribution of the schistosomiasis prevalence seems substantially higher in areas near the large dams, as shown in Figure 22. We are interested in whether the presence of a large dam affects the magnitude of the negative impact of schistosomiasis on agricultural yields: we therefore augment our initial model with an indicator variable $\mathbb{1}_{dam}$, representing whether a household is located in a province among those affected by the four large dams described in the paper, as well as with the interaction between schistosomiasis intensity and the indicator. The coefficient associated with $I_{jt} \times \mathbb{1}_{dam}$ represents the difference in the effect of schistosomiasis on log yield attributable to the presence of a dam. This specification allows us to disentangle the direct impact of dams, which should increase yields, from the deleterious indirect effect that they may produce by facilitating the diffusion of schistosomiasis. We note that we cannot have household fixed effects, as this estimation would not allow to estimate the coefficient of the (non-interacted) dam variable, and we rely on region fixed effects instead. The upper panel of Figure 23 reports the results: irrespective of estimation method, the interaction term is always statistically significant at conventional levels of confidence, and the presence of a large dam increases the average loss of agricultural yield due to schistosomiasis by a minimum of 8.2% to a

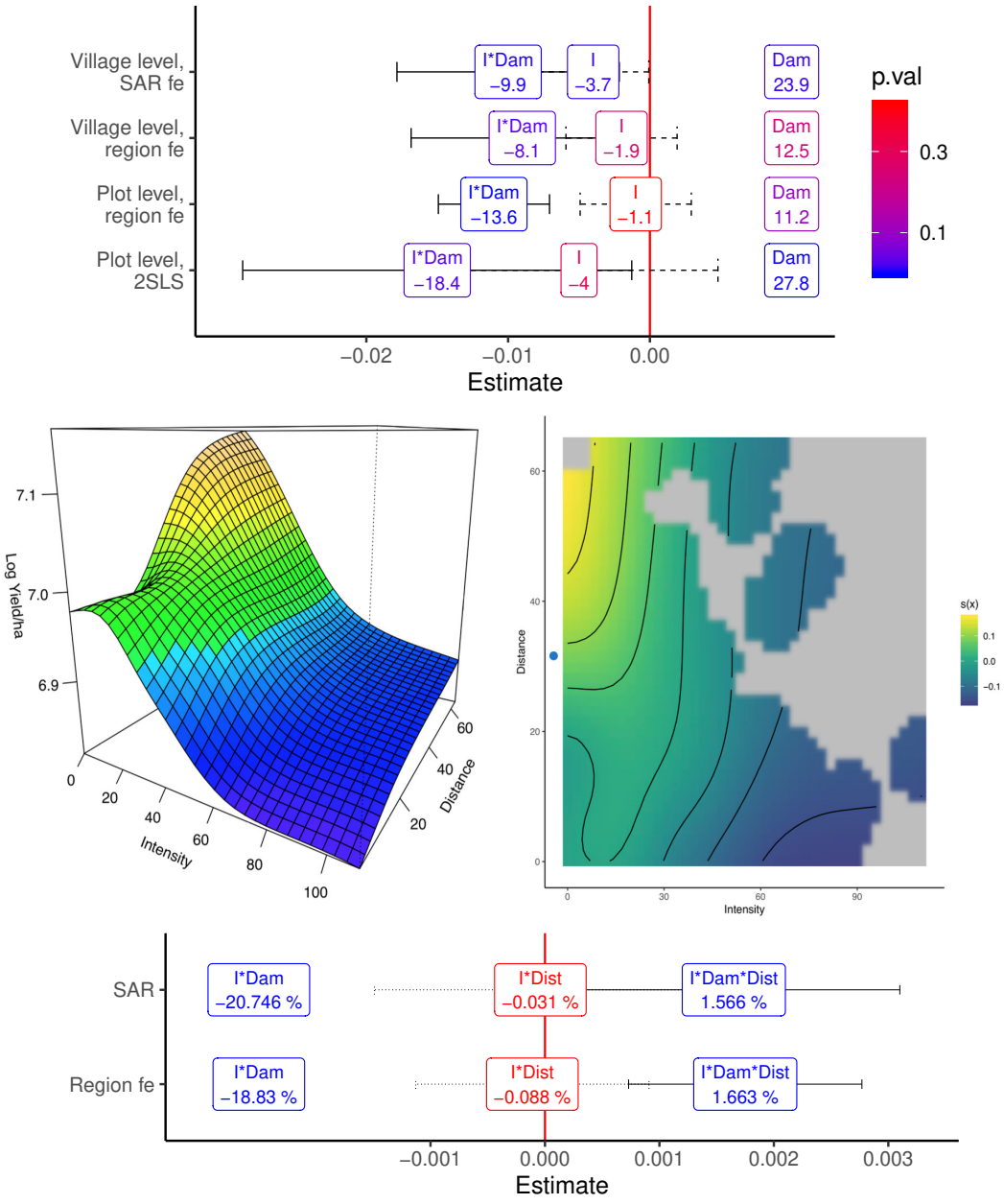

Figure 23: (Upper panel) Differential effect of schistosomiasis on yields caused by the presence of a large dam. (Middle panels) Estimated joint impact of schistosomiasis intensity and distance from all dams and water reservoirs. (Lower panel) Joint effect of schistosomiasis and distance from a large dam. All estimates control for time fixed effects, and plot-level estimates control for crop fixed effects. Errors are clustered at the regional level for region fixed effects estimates, and at the village level otherwise. 2SLS standard errors are cluster-bootstrapped.

maximum of 18.7%. The presence of a dam induces spatial effects that do not vanish, contrary to our earlier results: this is confirmed by how Moran's I statistic for the village-level model, even after controlling for region-specific effects, is still significantly positive (0.10, highly significant). Indeed, it is only when a SAR framework is adopted that the separate effects of intensity and dams are estimated with sufficient precision to make them statistically significant: the presence of a dam increases yield by 23.8%, and exacerbates the negative effect of schistosomiasis by 9.9%. Plot level estimation yields mixed results: the interaction term remains negative and precisely estimated, while the intensity variable becomes statistically indistinguishable from zero whether one controls for region fixed effects or applies the 2SLS procedure. The upshot is that our plot-level estimates suggest little effect of schistosomiasis outside of provinces with large dams, though our SAR results remain our preferred specification from the policy standpoint.

We refine the previous results by accounting for each village's distance (in km) from the nearest dam or water reservoir. We start by fitting an adaptive spline to the interaction of each village's distance from the dams and schistosomiasis intensity. The full network of dams and reservoirs is illustrated in Panel (b) of Figure 1 in the paper. Results are displayed in the middle panels of Figure 6: the left-hand panel shows how the deleterious marginal effect of schistosomiasis intensity on log yield is mitigated as one moves further away from a dam or a reservoir. The right-hand panel, based on the same estimation procedure, highlights how areas with high schistosomiasis intensity are concentrated within 20 km of a dam or reservoir. Households located in these areas suffer from large negative feedback effects between schistosomiasis and water resources development; to make matters worse, the effect of an increase in distance on the marginal effect of schistosomiasis intensity is greater for villages which display lower disease intensity. In order to find out what is the role of large dams in this mechanism, we include the triple interaction $I_{jt} \times \mathbb{1}_{dam} \times dist_j$, as well as all the double interactions, at a village level by both region fixed effects and SAR: the results are presented in the lower panel of Figure 6 and are very similar. The coefficient of the triple interaction is positive and significant: it has to be interpreted as the average amount to which the disease burden *decreases* as a village near any of the large dams moves one kilometer away from it. This implies an average *reduction* of around 1.5% (for SAR estimates) of the schistosomiasis burden on agricultural yield, starting from the average 18-20% loss suffered by the closest villages: being further away from large dams is therefore beneficial.

F Proof of the solution of the limiting disease SDE

The coupled SDE driving the disease dynamics is given by

$$dm_t = \left(\alpha(x_t)\frac{NBHm_t}{\beta + BHm_t} - \gamma m_t\right)dt + \sigma m_t dW_t.$$

We have shown in the paper that a measurable solution exists for the general case. We can solve explicitly for a limiting case: because of the assumptions on $\alpha(x)$ we have that when m_t is large $x \to 0$, and $\alpha(x) \to \alpha$. Furthermore, when m is large it's easily seen that the equation reduces to

$$dm_t = (\alpha N - \gamma m_t)dt + \sigma m_t dW_t.$$

which implies

$$dm_t - m_t(-\gamma dt + \sigma dW_t) = \alpha N dt.$$

The solution of this limiting SDE is obtained by entirely standard methods. Let us define an Itô process n_t such that

$$dn_t = n_1(t, n_t)dt + n_2(t, n_t)dW_t.$$

and n_1, n_2 are Lipschitz continuous functions. We would like this process to act as integrating factor for our SDE, that is

$$d(m_t n_t) = n_t(\alpha N dt),$$

implying

$$d(m_t n_t) = n_t dm_t - m_t n_t(-\gamma dt + \sigma dW_t). \tag{45}$$

Itô calculus allows us to write

$$\begin{aligned} d(m_t n_t) &= m_t dn_t + dm_t n_t + dm_t dn_t \\ &= m_t dn_t + m_t(n_1(t, n_t)dt + n_2(t, n_t)dW_t) + \sigma m_t n_2(t, n_t)dt. \end{aligned} \tag{46}$$

Using (45) we can obtain

$$-m_t n_t(-\gamma dt + \sigma dW_t) = m_t(n_1(t, n_t)dt + n_2(t, n_t)dW_t) + \sigma m_t n_2(t, n_t)dt.$$

Matching coefficients, we obtain the following relations:

$$\begin{aligned} -\sigma m_t n_t dW_t &= m_t n_2(t, n_t)dW_t \\ \gamma m_t n_t dt &= (m_t n_1(t, n_t) + \sigma m_t n_2(t, n_t))dt \end{aligned}$$

The first implies $n_2 = -\sigma n_t$, and plugging this in the second yields $n_1 = n_t(\gamma + \sigma^2)$.. This means that the integrating factor solves to

$$n_t = \exp\left(\int_0^t \left(\gamma + \frac{1}{2}\sigma^2\right) dt - \int_0^t \sigma dW_t\right).$$

when $n_0 = 1$, after applying Itô's lemma to $d \ln n_t$. Let us insert this result back in (46). We can write

$$\begin{aligned} d(m_t n_t) &= \alpha N n_t dt \\ m_t &= n_t^{-1}\left(m_0 + \alpha N \int_0^t n_s ds\right) \\ m_t &= \exp\left(-\int_0^t \left(\gamma + \frac{1}{2}\sigma^2\right) ds + \int_0^t dW_s\right) \times \\ &\quad \left(m_0 + \int_0^t \alpha N \exp\left(\int_0^s \left(\gamma + \frac{1}{2}\sigma^2\right) dr - \int_0^s \sigma dW_r\right) ds\right), \end{aligned}$$

which at the backward limit is (31).