Stichting Onderzoek Wereldvoedselvoorziening van de Vrije Universiteit

Centre for World Food Studies

Risk minimizing index functions for price-weather insurance, with application to rural Ghana

by

Michiel Keyzer
Vasco Molini
Bart van den Boom

Staff Working Paper **WP - 07 - 02** **May 2007**

Contents

Abstract

Poor farmers find it difficult to cope with price-weather shocks through self-insurance, because they cannot afford to keep large stocks and to protect their crops through irrigation and other measures. Mutual insurance is not an option either, because all participants would be faced with the same price-weather conditions at the same time. The next option of market insurance is plagued by excessive monitoring cost in avoidance of moral hazard and adverse selection. Consequently, new types of insurance are needed. Among the arrangements suggested, index-based insurance is currently receiving much attention. Index-based insurance offers an indemnification according to an index function that depends on agreed upon price and weather conditions rather than on an assessment of damage at individual farm level. Existing proposals and experiments present a synthetic index function whose effectiveness is established by assessing its capacity to stabilize revenues on the historical record. The present paper proposes an approach that is different in that it enables the insurer to offer an indemnification that is optimal from the perspective of the farmer in preventing a fall below a specified poverty line and is self-financing up to a given subsidy. To this effect, we develop and apply a model that minimizes farmers' risk of receiving an inadequate indemnification. The approach builds on methods from catastrophic risk management in insurance and support vector regression in statistics. It is applied to Ghana, where according to our database, 47 percent of the farm population fell below the poverty line. Simulations of index-based insurance show that, while a parametric form fits reasonably the ideal indemnification it reduces poverty by only 4 per cent. By contrast, the proposed semi-parametric forms perform much better in terms of fit and could reduce poverty by another 5-10 per cent points, depending on the regularization.

1. Introduction[1]

For most farmers in Sub-Saharan Africa (SSA), yields are generally low and subject to variation due to erratic rainfall patterns and frequent outbreaks of pests and diseases. Consequently, households live with the permanent threat of crop failure (Dercon, 1996). Moreover, the prices of major tropical crops produced in the area tend to fluctuate widely from year to year, following world prices on markets where supply variations face an inelastic demand and obeying patterns that are largely independent from local crop conditions, and, therefore, do not compensate for poor harvests (Fafchamps, 1992).

Farmers address these shocks in various ways. One is self-insurance, whereby they take various measures in the physical sphere, such as keeping cash crops in stock until the price recovers, depleting food inventories in dry periods, irrigating dry fields, slaughtering cattle, and selling off jewels. Clearly, these options are hardly within reach of poor farmers, who would have to choose between eating the seeds of the next planting season and selling off their last asset. Futhermore, self-insurance may also reduce the capacity of households to escape from poverty as farmers engage themselves in safer but on average less remunerative activities. In a study on Zimbabwe, Elbers et al (2007) find a drop of 46 percent in capital accumulation.

Mutual insurance is not a promising option either when it comes to dealing with price-weather shocks, because all farmers in the collective are affected by the same price and weather adversities at the same time. Market insurance is on its part plagued by excessive monitoring costs of measuring damages. Furthermore, the insured farmer may not pay the premium due, or do this too late. Finally, the insurance company itself might default on its obligations. Because of this high cost of transactions and the inevitable moral hazard, poor farmers often find themselves trapped in poverty, unable to take advantage of upcoming profitable opportunities, because the risk is too high for them to take a chance.

This situation has become a major concern of the development community, particularly at a time that arrangements to control the markets directly, through prices and stock management, have been discarded as less effective. There is a growing literature on pathways to insure the poor better, arguing that a broadening of saftety nets is warranted, not only for equity reasons, but increasingly so to overcome welfare losses and poverty traps caused by farmers' response to uninsured risk (Dercon, 2005). In urban areas, instruments such as, health insurance and free provision of drugs have been introduced for this purpose. However, in rural areas, where few public facilities are available, more general arrangements such as crop insurance, possibly subsidized, seem preferable.

As compared to purely publicly funded schemes, subsidized crop insurance offers the advantage that the insured groups pay some premium themselves. Besides alleviating the pressures on national budgets, this would reduce free riding, since participants who pay their dues will tend to exercise some countervailing pressure when they see that others benefit more than they deserve contractually.

Also, by creating pools of policy holders with different risk profiles, say, of farmers with different cropping patterns, crop insurance offers possibilities for mutual insurance at above village level, and also for solidarity between population groups, all this of course within the limits set by political realities. Furthermore, at international level, market insurance is considered an interesting option, because the economies in rural SSA are small by international standards,

[1] Paper written for the project "Can Weather and Price Insurance Reduce Rural Poverty in Africa?". The authors want to acknowledge support from The World Bank (Research Grant no. 7131322) and from the Ghana Statistical Service that kindly provided access to the data, assisted in compiling the price series and organized a seminar in July 2006 to present the work to the public in Ghana. Particular thanks go to the Acting Government Statistician, Dr. N.N.N. Nsowah-Nuamah. Also, comments by Dr. Ernest Aryetee, Dr. Kwadwo Asenso-Okyere, Dr. Wouter Zant and participants to the Oxford's CSAE 2007 conference are gratefully acknowledged.

which makes reinsurance relatively easy. Finally, even when subsidized, crop insurance arrangements are considered less distortionary than many other farm support measures and so far enjoy a Green Box status in the WTO.

However, the issue has to be addressed that crop insurance suffers from the classic problems of adverse selection and moral hazard, and, as mentioned, has high costs of administration and monitoring. In this connection, index-based insurance has been proposed as an adaptation that is better tailored to developing country circumstances since it reduces moral hazard and adverse selection and lowers administrative and monitoring costs (Skees et al. 1999; Skees et al. 2005). The idea is to condition indemnity payments on variables that are independent from both farmers and insurers decisions but sufficiently correlated with farm income, nonetheless.

Index-based insurance pays out when an agreed upon indicator falls below an agreed upon threshold. Classical examples are payments triggered by the recorded rainfall at a weather station or, by the price at a local market, or by weather-price conditions simultaneously. Reviews of revenue insurance (Mahul and Wright, 2003) indicate, however, that it is not easy to piece together by synthetic means a function that predicts well actual individual damages. Yet, if a function with this property could be designed, index-based insurance becomes an attractive alternative for traditional crop insurance and has an enhanced scope for dealing with covariate risks through reinsurance.

Nonetheless, index-based insurance necessarily leaves part of the risk (basis risk) with the farmer, due to the anchoring on a limited number of variables, the fact that contracts are not individualized and quantified on the basis of past data. This basis risk is particularly worrisome as situations may occur where payments predicted by the index grossly underestimates actual damage (Goodwin and Mahul, 2004; Barnett et al., 2006). It will also be higher when the indemnification schedule is poorly adapted to the needs of policy holders, which is likely to be the case for functions that are postulated on a priori grounds and applied to a collective of farmers with varying risk profiles.

Indeed, there is a danger of the lower implementation cost to come with an upsurge of basis risk. Therefore farmers' willingness to buy index-based insurance will naturally be contingent on the specific design of the contract, in particular on whether indemnity payments correlate sufficiently high with individual damages, and on modalities of premium payments.

Index-based insurance is now piloted in several developing countries, including Mexico (Skees et al., 2001, 2005), India (Kalavakonda and Mahul, 2005; Veeramani et al., 2005; Zant, 2005), Ghana (Sarris, 2002) and Malawi (Hess and Syroka, 2005). Most applications are still under study, in a pilot phase, or subject to revision but experience gained until present would seem to indicate that poor farmers' remain reluctant to buy index-based insurance, despite significant subsidies often offered on the premiums. One reason might be that based on the indemnifications paid so far, they consider the basis risk relatively high under the proposed schemes, another that they do not see how the current participation by their neighbours could ever overcome the problems of covariate risk they so often faced under mutual insurance arrangements. Hence our attempt to design an insurance that minimizes basis risk and can be specified for any pool of farmers willing to share risk and for various levels of self-financing.

Whereas index-based insurance arrangements proposed so far offer indemnification according to some a priori postulated function whose performance is being assessed, we present a methodology and an application to design schedules that are optimally adapted to specified needs of participating farm groups and that is self-financing up to a specified subsidy. It will be indexed on few, relatively easily observable variables — prices, rainfall and farmsize — giving up the option of full adaptation to individual needs. One reason is that the policy uniformly applies to all policy holders with many individuals having the same observed prices and weather, about the same farmsize but, say, different cropping patterns, yields, and personal situations. An index function will, whatever its shape, only compute a single indemnification for all of them. The second reason is that the function has to apply in the future, under circumstances other than those found in the historical record. Therefore, it is important to avoid overfitting whereby the function may nicely fit the past but perform poorly under conditions beyond the historical record.

To deal with both aspects, we adapt semi-parametric techniques of Support-Vector (SV) regression (Vapnik, 1998; Schoelkopf and Smola, 2002; Herbrich, 2002). The procedure optimally fits to the needs of individual farmers in that it minimizes their basis risk within a self-financing framework and deals with overfitting by extending the estimation problem with out of sample constraints, and by a penalization of flexibility (regularization).

The remainder of the paper is organized as follows. Section 2 deals with the theory. Section 3 discusses computation of the ideal individualized indemnification. In Section 4 we estimate indemnification functions that are as close as possible to this ideal. Section 5 concludes.

2. Theory: optimal indemnification

Optimal indemnification of a single policy holder under catastrophic risk

We start generating insurance policies with indemnification schedules derived from risk minimizing approaches under catastrophic risk. The idea is that these schedules should avoid ruin for all poor farmers in a particular population group, under assumed risk pooling with richer farmers.

Indemnification inclusive per capita income $r(z,\tau,\varepsilon)$ is expressed as function of decisions $z \in Z \subset R^m$, where Z is a nonempty, compact and convex set, of premium $\tau \in R_+$, and of events $\varepsilon \in R^n$ described by density g. Ruin is expressed as the income threshold $\underline{r}$ below which poverty becomes extreme. Decisions z comprise the entire risk management strategy of the farm household, inclusive of self- and mutual insurance as well as commercial insurance, if available, but exclusive of premium τ, which farmers must take as given. As in Ermoliev et al. (2000), the agent maximizes expected utility, while trying to avoid disaster:

$$max_{z\in Z} Eu(r(z,\tau,\varepsilon)) + \rho E\, min(u(r(z,\tau,\varepsilon) - u(\underline{r}),0), \tag{2.1}$$

where u is a concave increasing utility function, expectation E is taken with respect to ε, and ρ is a given and high penalty coefficient, reflecting that the household seeks to avoid ruin by all means. Therefore, we may when considering the optimal insurance limit attention to the minimization of expected income shortfalls:

$$max_{z\in Z} E\, min(u(r(z,\tau,\varepsilon) - u(\underline{r}),0), \tag{2.2}$$

We treat the purchase of insurance as a discrete risk management strategy $z = 1$, and henceforth drop choice variable *z*. This is possible because we are concerned with the design of an ideal insurance that operates as last resort and is unrestricted by the payoff structure of any particular asset. Consequently, when providing such insurance, we may consider all other risk coping decisions to have been made already.

Insurance provides access to an indemnification profile, denoted by $v^+(\tau,\varepsilon)$. For given poverty line $\underline{r}$, income profile of income $h(\varepsilon)$ before indemnification, $0 \le h(\varepsilon) \le \overline{h}$, and premium τ, the indemnification needed to supplement farmer's income in case of events ε that bringing his income below the poverty line equals:

$$v^+(\tau,\varepsilon) = max(\underline{r} + \tau - h(\varepsilon), 0). \tag{2.3}$$

Under this arrangement, insured income:

$$r(\tau,\varepsilon) = h(\varepsilon) + v^+(\tau,\varepsilon) - \tau \tag{2.4}$$

will never fall below poverty line $\underline{r}$, whereby expected shortfall (2.2) will be zero. However, this is for given premium. To keep the arrangement self-financing, the premium has to be such that the deficit:

$$F(\tau) = \int v^+(\tau,\varepsilon)g(\varepsilon)d\varepsilon - \tau \tag{2.5}$$

is exactly covered by a (net) subsidy, denoted σ and covering possible implementation cost as well as part of the payments:

$$F(\tau) = \sigma\,. \tag{2.6}$$

We note that, for given subsidy, it might be impossible to meet this self-financing constraint. After all, the insurance can be no more than a smoothing device, implying that the subsidy would need to satisfy the following feasibility requirements.

ASSUMPTION 1 (feasibility of self-financing insurance that eliminates poverty):
(i) Subsidy σ is sufficient to cover the shortfall of expected income below the poverty line:

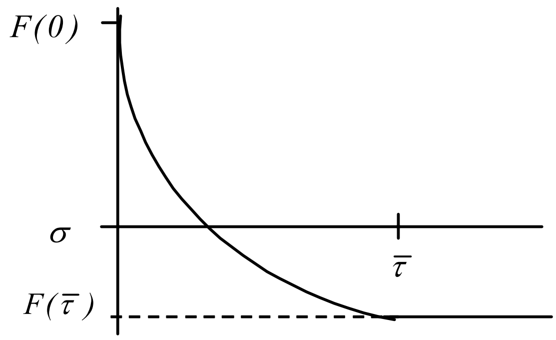
$$\sigma \geq \underline{r} - \int h(\varepsilon)g(\varepsilon)d\varepsilon\,.$$

(ii) Subsidy σ does not surpass the total need for indemnification:

$$\sigma \leq \int max(\underline{r} - h(\varepsilon), 0)g(\varepsilon)d\varepsilon\,.$$

Note that both conditions are irrespective of premium τ. Indeed, since revenue $r(\tau,\varepsilon)$ is nonincreasing in τ for every ε and decreasing until the premium reaches its maximum $\overline{\tau} = max(\overline{h} - \underline{r}, 0)$, it follows that $F(\tau)$ is decreasing . Also, since $F(\overline{\tau}) \leq \sigma$, by Assumption 1.(i), while $F(0) \geq \sigma$ by Assumption 1.(ii), the solution $F(\tau) = \sigma$ exists and is unique. This is illustrated in Figure 1 at the point where downward sloping curve $F(\tau)$ intersects with horizontal line σ.

FIGURE 1. Self-financing and subsidy of an ideal insurance against poverty



Risk pooling

Next, we allow for solidarity through risk pooling among policy holders, still dealing with ideal insurance in the sense that every farmer holds a fully individualized contract. To represent this, we distinguish groups indexed i, consisting of N_i individuals with per capita income profile $0 \leq h_i(\varepsilon) \leq \overline{h}_i$, poverty line $\underline{r}_i$, and premium τ_i. We allow for premium differentiation across groups through flat premium per hectare τ with corresponding per capita premium $\tau_i = \gamma_i \tau$ in accordance with per capita farmsize $\gamma_i > 0$. Indemnification is now defined as:

$$v_i^+(\tau,\varepsilon) = max(\underline{r}_i - h_i(\varepsilon) + \gamma_i \tau, 0), \tag{2.7}$$

leading to a budget deficit:

$$F(\tau) = \sum_i n_i \left(\int v_i^+(\tau,\varepsilon) g(\varepsilon) d\varepsilon - \gamma_i \tau \right), \tag{2.8}$$

where $n_i = N_i / \sum_i N_i$ is the share of group i in total population. This deficit is to be covered from a subsidy, as in (2.6). The generalisation of Assumption 1 reads:

ASSUMPTION 2 (feasibility of self-financing risk pooling insurance that eliminates poverty):
(i) Subsidy σ is sufficient to cover the shortfall of expected income below the poverty line:

$$\sigma \geq \sum_i n_i \left(\underline{r}_i - \int h_i(\varepsilon) g(\varepsilon) d\varepsilon \right).$$

(ii) Subsidy σ does not surpass the need for indemnification:

$$\sigma \leq \sum_i n_i \int max[\underline{r}_i - h_i(\varepsilon), 0] g(\varepsilon) d\varepsilon .$$

After replacing (2.5) by (2.8), under Assumption 2, self-financing condition (2.6) identifies a unique premium. Hence, a group of farmers who pool their risk could steer clear of poverty by contributing a common premium per hectare, which, in combination with an externally provided per capita subsidy, can finance all claims of those whose income falls below the poverty line.

Discrete states of nature

Next, to compute the ideal indemnification to which estimation will seek adaptation, we shift from continuous density to a discrete representation by an empirical distribution with states of nature indexed $\ell = 1,...,L$, inferred from the historical record, with expectations approximated as means over these states.

The discrete representation of (2.8)-(2.6) considers state-specific per capita income $h_{i\ell}$, indemnifications $v_{i\ell}^+(\tau) = max(\underline{r}_i - h_{i\ell} + \gamma_i \tau, 0)$, and mean deficit $F(\tau) = \frac{1}{L} \sum_i n_i \left(\sum_\ell v_{i\ell}^+(\tau) - \gamma_i \tau \right)$. Letting $\kappa = \sum_i n_i \gamma_i > 0$ denote mean per capita farmsize, the ideal insurance can be obtained from the linear program:

$$\begin{aligned}
& min_{\tau, v_{i\ell}^+, \zeta_{i\ell} \geq 0} \quad \frac{1}{L} \sum_i n_i \sum_s \zeta_{i\ell} \\
& subject\ to \\
& \quad \zeta_{i\ell} \geq \underline{r}_i - h_{i\ell} - (v_{i\ell}^+ - \gamma_i \tau) \\
& \frac{1}{L} \sum_i n_i \sum_\ell v_{i\ell}^+ = \kappa \tau + \sigma
\end{aligned} \tag{2.9}$$

Indeed, under Assumption 2, this program will find a unique optimal premium and an indemnification profile with zero risk $\zeta_{i\ell} = 0$, i.e. build the ideal asset against poverty over the given sample.

Optimal index-based insurance

Given ideally individualized per capita indemnifications $v^+_{i\ell}$ obtained from (2.9), our index-based insurance seeks to obtain some function $f(x)$ that fits these indemnifications on some vector x of variables appearing in the index. However, as this fitting is a risk minimizing operation in its own right, both choices are looked at jointly, to account for tradeoffs and priorities.

At this point, it may be recalled from the introduction that index-based insurance will never be able to eliminate all risk, first because it is represented by a non-individualized function that, whatever its shape, can only compute a single indemnification for all participants of a given farmsize who face the same weather and price conditions, and second because the function has to apply in an uncertain future, under circumstances other than those found in the historical record.

This also suggests that statistical estimation would seem the natural way to proceed but we need to choose an estimation technique that is especially flexible, because unlike in the common situation in econometrics, the set of explanatory variables at hand is limited, and since individualized contracts are ruled out, the possibility of keeping track of fixed effects by household is ruled out. At the same time, again because the contract is to apply in the future, the pitfall must be avoided of overfitting to past observations. As estimation technique, we therefore opt for semi-parametric Support-Vector regression. SV-regression has capacity to adjust flexibly to any data set while addressing the issue of overfitting, and it can naturally be merged with the risk-minimizing framework specified above.

We remark that since under the national risk pool we are not going to allow for individualized contracts, this national arrangement considers every outcome of group i an event. that is a possible outcome for any individual. Hence, events comprise all states ℓ for each group *i*. For convenience, we denote these by $s = 1,...,S$ with $S = I \cdot L$ and denote group *i* associated to *s* by i_s. Skipping technicalities, we admit the following class of functions, in accordance with the Representer Theorem (Kimeldorf and Wahba, 1970):

$$f(x) = \sum_s \alpha_s k(x_s, x) + \sum_j \beta_j \phi_j(x), \tag{2.10}$$

where *k* is a given kernel function and ϕ_j are given functions as well. The first term on the right hand side is the non-parametric term, the second the parametric one.

We remark that variables x entering the index will vary across participants only to the extent they face different weather conditions and prices. In addition, household-specific events such as illness could also be included, provided a credible assessment of their occurrence is agreed upon.

Optimality of the proposed insurance will be expressed in terms of minimum risk, where the risk consists of both income shortfalls, as in (2.9), and fitting errors. In addition, the risk criterion is extended by a quadratic regularization term $\frac{1}{2}\sum_s \sum_r \alpha_s \alpha_r k(x_s, x_r)$ multiplied by a scalar λ_S. This term penalizes coefficient values and is essential to avoid overfitting and can be made to shrink as sample size increases, and to vanish asymptotically. After putting factor $\rho > 0$

on the priority to avoidance of income shortfall, as in basic program (2.1) for avoidance of catastrophic risk, we can write the regularized risk minimizing problem as the quadratic program:

$$\begin{aligned}
&\min_{\tau, v_s^+, \zeta_s \geq 0; \xi_s, \xi_s^* \geq 0, \alpha_s, \beta_j} \rho \frac{1}{L} \sum_s n_{i_s} \zeta_s + \frac{1}{L} \sum_s n_{i_s} (\xi_s + \xi_s^*) + \frac{\lambda_S}{2} \sum_s \sum_r \alpha_s \alpha_r k(x_s, x_r) \\
&\text{subject to} \\
&\quad \zeta_s \geq \underline{r}_{i_s} - h_s - (v_s^+ - \gamma_{i_s} \tau) \\
&\quad \frac{1}{L} \sum_s n_{i_s} v_s^+ = \kappa\tau + \sigma \\
&\quad v_s^+ = \sum_r \alpha_r k(x_r, x_s) + \sum_j \beta_j \phi_j(x_s) + \xi_s - \xi_s^* \\
&\quad \frac{1}{L} \sum_s n_{i_s} [\sum_r \alpha_r k(x_r, x_s) + \sum_j \beta_j \phi_j(x_s)] = \kappa\tau + \sigma
\end{aligned} \tag{2.11}$$

Next, we recall from (2.1) that for sufficiently large factor ρ, we can under Assumption 2, solve (2.9) ahead of (2.11), treating optimal premium τ and indemnification profile v_s^+ as given data in (2.11), and limit attention to estimation of parameters (α, β). Next, writing y_s for v_s^+, defining mean resources $\bar{\sigma} = \kappa\tau + \sigma$, and allowing for soft margin η (with given penalty factor ϑ), we obtain an SV-regression problem, incremented with a financing constraint:

$$\begin{aligned}
&\min_{\xi_s, \eta \geq 0, \alpha_s, \beta_j} \frac{1}{L} \sum_s n_{i_s} \xi_s + \frac{\lambda_S}{2} \sum_s \sum_r \alpha_s \alpha_r k(x_s, x_r) + \vartheta\eta \\
&\text{subject to} \\
&\quad y_s \leq \sum_r \alpha_r k(x_r, x_s) + \sum_j \beta_j \phi_j(x_s) + \xi_s + \eta \\
&\quad y_s \geq \sum_r \alpha_r k(x_r, x_s) + \sum_j \beta_j \phi_j(x_s) - \xi_s - \eta \\
&\quad \frac{1}{L} \sum_s n_{i_s} [\sum_r \alpha_r k(x_r, x_s) + \sum_j \beta_j \phi_j(x_s)] = \bar{\sigma}
\end{aligned} \tag{2.12}$$

Note that we replaced the indemnity constraint in (2.11) by two inequality constraints, which are never effective simultaneously. Hence, we can work with a single vector ξ of absolute value deviations to which soft margin η is added. SV-regression in fact owes its name to this margin, which is introduced to avoid excessive focus on small errors (Huber-robustness, Huber, 1981). Together, the regularization term and the margin prevent each observation from demanding a large non-zero value for its own parameter.

We note that the parametric term could be used to represent an a priori given schedule of indemnification. This could serve as starting point, with the kernel terms measuring the inadequacy (or correction) of this schedule, whereas the soft margin and the estimation error reflect the basis risk that remains with the farmers, which, for large samples will approximate the risk that is unavoidable for the given anchors of the insurance.

Program (2.12) differs from the standard form of SV-regression because of the additional self-financing constraint. This possibility of inserting such a constraint is essential for index-based insurance, since it makes it possible to ensure that the proposed index will satisfy financing requirements. Furthermore, other constraints can be imposed as well. First, solvency constraints would seem of relevance, because the financing constraint in (2.12) only requires the insurer to meet the contractual obligations in the mean, neglecting the fact that the insurer should be able to pay every year from the start of the arrangement, also in case of an initial period of adversity. Thus, in its present form program (2.12) assumes that the arrangement enjoys a public guarantee, either from national government or from international donors, exempting it from solvency restrictions, which could though be incorporated as limits on the cumulative payments over specified sub-samples. Likewise, restrictions could be introduced to target subsidies in favour of relatively poor groups or to limit net contributions of relatively rich groups.

From a practical perspective, the key feature of (2.12) is that all constraints are linear in choice variables and that the objective is linear-quadratic and convex. This enables us to solve this program numerically by standard tools of quadratic programming, as is also usual in SV-regression, because all k and ϕ terms can be evaluated ahead of the program leaving expressions in α, β, η and ξ only.

Finally, regarding the likely performance of the arrangement in the future on new data not accounted for in the estimation, the technique's capacity to learn from past events is essential. As this is a rather technical issue, we only mention two properties inherited from SV-regression. The first is that minimization in (2.12) of the sum of absolute values of errors amounts to estimating the conditional median (Koenker and Bassett, 1978), as opposed to the conditional mean estimated by least-squares methods, which makes it less sensitive to outliers. The second property is that for $S \to \infty$ the estimate converges strongly to the true conditional median under appropriate reduction of the regularization factor and the soft margin (Takeuchi et al., 2005; Norkin and Keyzer, 2007).

3. Application To Ghana: Ideal Individualized Insurance

Data compilation

We now proceed with an application for Ghana. This application starts with the compilation of a database of representative farm households for 26 different states of the world that reflect economic and weather conditions as derived from the historical record. We treat them as independent and identically distributed observations from a stationary distribution.

The data sources include the Ghana Living Standards Survey GLSS, covering the years 1987/88, 1988/89, 1991/92 and 1998/99 and the Population Census of 1970, 1984 and 2000 (GSS, 1987/2001; GSS, 1989/95/2000). In addition, for the period 1980 to 2005, time series data are available for monthly rainfall data at 40 stations throughout Ghana and of Accra-prices for all main crops (GMI, 2006; GSS 2005). To obtain the income profile of each representative household under the respective states of the world, we have derived from the four rounds 50 groups with fixed size of land holding that show, on average, homogeneous characteristics in terms of expenditure patterns and income structure across rounds. Yet, several limitations of the data set came up in the process.

One is that division of crop production by area cultivated often gave yields that conflicted seriously with agronomic information (FAO, 2005). This might be because farmers did not correctly report about their landuse. Such an inconsistency between the production section and land section of the questionnaires was reported earlier for cocoa (Teal and Vigneri, 2004) but it appears to be even more common for the other crops. In addition, data on full income (agricultural plus non-agricultural income) appeared to be poorly correlated with expenditures. This is presumably due to the general underestimation of the per capita rural expenditures in the survey (Fofack, 2000).

At the same time, land per capita proves to be highly correlated with expenditures in all survey rounds and its distribution relates well to the expenditures distribution. Furthermore, land holding size per capita seems to offer a good identifier when constructing a quasi-panel (Deaton, 1985; Beaudry and Green, 2000) over the four rounds since, unlike income and expenditures classes, holding size classes tend to vary little across the four rounds and to correspond to relatively stable income and expenditure patterns. It also offers a more direct link than income distribution to the resource endowments and geographic location of farmers (Decaluve et al., 2001). Hence, we grouped households in quintiles of land holding size per capita for each of Ghana's 10 regions, arriving at 50 representative agents (groups). Next, to obtain the time series of per capita income of each group, we proceeded as follows.

Regarding income from farming, we computed by region for every year the total agricultural farm income based on agricultural production and price observations, and distributed these across farm groups in proportion to their land holding size. To calculate yields and crop production we extracted information on the cropping pattern from survey rounds. To obtain yield in the different states of the world, we constructed time series of regional yields from climate and agronomic data (Chapagain and Hoekstra, 2004). These data were integrated with the information on yield variability extrapolated from surveys in order to maintain realistic differentiation in cropping patterns between groups.

Regarding non-crop income, we added the non-crop income by group, as recorded in the the four survey rounds, with an upward correction in case total income fell below total expenditures in the survey. Non-crop income is is essential as it enables us to account for all risk coping actions undertaken by farmers, using sources external to their main activity, through remittances, seasonal jobs outside agriculture, retail trade activities etc. Non-crop income makes it possible to assume that the actual income profile is inclusive of all other risk management strategies such as joining a mutual insurance, irrigating, and modifying the crop composition. Indeed, as discussed in relation to ideal asset (2.2), the insurance can be interpreted as a proxy of an ideal financial asset precisely because it is conditional on farmers' risk management strategies in all other domains, real as well as financial.

In this way, household groups with characteristics invariant across the 26 states of world were constructed as representative agents for the decennium 1987-1998 but clearly, the transition from around 2000 farm households in the surveys to 50 representative agents amounts to a reduction in variability within groups. We account for this by maintaining major information on expenditures distribution within regions and quintiles, in two ways. First, we further subdivide the groups in two sub-groups say, one relatively poor, endowed with an amount of land per capita closer to lower bound of every quintile, and one richer, with an amount closer to the upper bound. This discards all other distinctive features across households within each quintile and region, but as discussed, maintaining differences in land holding size already keeps track of the major indicator of disparity.

Second, given this split with corresponding population fractions P_G of the poor, the per capita farmsize of the two subgroups is obtained as:

$$\begin{aligned} \gamma_{PG} &= (1-P_G)\gamma_{LG} + P_G\gamma_{MG} \\ \gamma_{RG} &= P_G\gamma_{UG} + (1-P_G)\gamma_{MG} \ , \end{aligned} \tag{3.1}$$

where subscripts *L*, *M* and *U* refer to lower, median and upper bound of every group. To the new representative agents is then attributed a new population weight. Finally, the poverty line of each subgroup is adjusted to reproduce the income shortfall of that group as measured in the original data set. Hence, our calculation maintains some of the measured income disparity within groups, and reproduces the observed income shortfall below the poverty line of the group. Summary statistics are reported in Appendix I.

Ideal individualized indemnification for three types of insurance pools

To be of interest to farmers, an indemnification schedule should on the one hand be kept sufficiently flexible to fit their needs, and on the other hand apply to a pool of farmers who are willing to share their risk. Consequently, to succeed any initiative to introduce insurance will have to build on existing expressions of solidarity as reflected in prevailing mutual arrangements.

In this section, we restrict attention to three hypothetical insurance pools and compute the ideal insurance needed to avoid all income shortfalls below the poverty line in the pseudo-panel constructed for Ghana, in accordance with Figure 1 and program (2.9), for different financing constraints expressing the participation of groups i in the arrangement. The per hectare premium τ will solve (2.6) and is self-financing up to a given external subsidy σ. To take into account potential implementation costs and to attract insurance companies, we augment the premium with a mark up of around 8 percent, as in Hess and Sykora (2005).

In any case, the design of an ideal insurance requires average income in a pool plus the subsidy to equal at least the insured minimum income (Assumption 2). Consequently, an insurance against poverty for farmers who are structurally below the poverty line can only work in a pool that also comprises relatively rich farmers, or requires heavy financial support. The three arrangements considered are reported in Table 1.

TABLE 1. Per hectare premium under group pool, regional pool and national pool

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Regional scheme	National scheme
Ashanti	1441	504	258	73	3	127	265
Brong-Ahafo	632*	590*	799*	152	9	161	265
Central	1100*	1011	379	92	16	147	265
Eastern	968*	1336*	299	84	5	280	265
Greater Accra	1393	170	26	61	24	154	265
Northern	1359*	578*	410*	545*	48	876*	265
Upper East	1585*	694*	754*	791*	45	1187*	265
Upper West	1543*	584*	467*	633*	53	857*	265
Volta	508*	914*	817	103	6	650	265
Western	736	535	144	16	2	34	265

*Asterisks point to the presence of a subsidy per capita on the premium paid.
**The regional subsidies amount to 112.000 cedi per capita in Upper East, 65.000 in Northern and 96.000 in Upper West region.
***The group subsidies range from zero and small values in the upper quintiles to as much as 300.000 in the first quintile of the poorest regions.
Source: GSS (1988/89/92/98) and authors' calculations.

a. Full solidarity.

The pool consists of all farm households in the country, in this case no subsidy is required (column 7). The arrangement would operate as a social security fund financed from a fixed land tax. The fund would pay out to those who would otherwise fall into poverty. The self-financing ensures that the taxation exactly suffices for the expected indemnification. The annual premium is around 237,000 cedi per hectare, or about 12% of estimated average per hectare income of 1,965,000 cedi in the period considered.

b. Regional solidarity.

The pool includes all farmers living in the same region. Program (2.9) is now solved for each of the 10 regions, given a region-specific net subsidy σ_R The premium varies strongly across regions (column 6). For example, for the three regions in the North, average income of all farmers is below the poverty line of 700,000 cedi per capita per year. In such a situation, it clearly is impossible to cover the risk of falling into poverty, and a minimum subsidy of about 10% of the average income is required to bridge the gap. While the South has only few cases of income shortfalls, almost all confined to the first two quintiles, in northern regions the frequent shortfalls imply that significant subsidies are needed on top of the premium.

c. *Group pooling.*

The pool includes all farmers living in the same region and belonging to the same land per capita quintile. The arrangement now has a group-specific per capita net subsidy σ_G and mainly affects the income distribution in the lower quintiles. In the first quintiles, all farmers are lifted above the poverty line thanks to the highly subsidized indemnifications (columns 1 and 2). From the 3rd quintile upward, excluding the three Northern regions, the indemnification by group is sustainable without external funds.

4. Index-Based Insurance: Estimation And Simulation

Specification

The present section reports on the estimation of index-based insurance schedules designed through semi-parametric regression of the ideal indemnification discussed above on weather-price variables and farmsize, following program (2.12). The software package described in Keyzer (2005) is used for computation. Since the aim of the exercise is also to compare our approach with the synthetic forms in the literature, we also discuss results from fully parametric regressions, which can be seen as intermediate in terms of flexibility between synthetic and semi-parametric forms.

The data comprise $I = 100$ households under $L = 26$ states of nature, leading to a sample size of $S = I \cdot L = 2600$. As index variables we use the Length-of-Growing-Period (x_1), six prices (x_{2-7}) covering both cash and staple crops (cocoa, cassava, yam, cocoyam, maize and tomatoes), and per capita farmsize x_8. The parametric form is postulated to be linear $\phi_j(x) = x_j$ with an additional $\phi_0(x) = 1$ for the constant.

The estimation proceeds in three steps, as in the back-fitting procedure described in Schoelkopf and Smola (2002): (i) estimate the parametric part with coefficients β; (ii) keeping β fixed, estimate the coefficients α of the non-parametric part; (iii) joint estimation of α and, as in (2.12).

To estimate purely parametric index $\sum_j \beta_j \phi_j(x)$, we implement program (2.12) keeping $\alpha = 0$, or, equivalently, taking the regularization factor so high that the non-parametric part phases out. The program then defines a weighted Least-Absolute-Deviation (LAD) estimator (Gilonia et al., 2006), but extended with financing constraint (2.6), and with a provision for a soft margin, that decomposes the error into a common term, the η-margin that avoids penalization of indistinguishable observations within a band[2], and the remaining idiosyncratic error.

Turning to non-parametric part $\sum_r \alpha_r k(x_r, x)$, we estimate α in (2.12), this time keeping β fixed. We use the Gaussian kernel with a window width that is one tenth of the one that is optimal under Normally distributed samples (Haerdle, 1995). This is done to keep program (2.12) tractable in size, at the expense of a reduced capacity of accounting for interdependencies in the data. Yet, although reduced in number, the remaining interdependencies show a meaningful pattern, maintained nonzero kernel terms among sites with similar rainfall pattern and similar land holding size. By contrast, extending the window size above the chosen value tends to overstate the interdependencies, allowing for very different circumstances to co-determine a single indemnification.

Recalling that the level at which regularization factor λ is kept acts central lever to modulate the performance of the semi-parametric regression, we scan over various λ-values, starting from zero upwards to find the best value. At $\lambda = 0$, we have over-fitting and maximal fit inside the sample, but the nonparametric part becomes "bumpy" with large positive as well as negative α-values, which foretells poor out-of-sample performance. At the other extreme, $\lambda = \infty$, we return to purely parametric regression. We eventually select a λ-value that is sufficiently high to reduce substantially the variability and the number of nonzero α_s on the one

[2] In fact, the computations use a fixed instead of a soft margin, set at $\eta = 20{,}000$. The Lagrange multiplier indicated that this level amounts to a penalization factor near unity.

hand, and not too high to loose the flexibility of semi-parametric function (2.10) on the other. Specifically, the path for scanning is specified as: $\lambda(1)=0$; $\lambda(2)=\delta$; $\lambda(n)=\delta 2^{n-2}$, $for\ n=3,4,5,6$.

Results from estimation

Table 2 shows the parameter estimates of a purely parametric index-function (putting α to zero), and Table 3 the results for the semi-parametric index using the back fitting procedure (keeping β fixed). In Table 4, after having selected the optimal λ -value, we report on results of joint estimation of (α , β). The three tables refer to the index-based insurance of the national pool only (Table 1, last column). Results for contracts at the region- and the quintile- level are qualitatively similar, with noted differences with respect to the level of self-financing (see footnotes Table 1) and a more pronounced tendency for income distributions to collapse near the poverty line.

TABLE 2. Estimated parametric index function ($\alpha=0$, $\beta_0=638$, $R^2=0.54$)

β	LGP	Cocoa	Cassava	Yam	Cocoyam	Maize	Tomatoes	Farm size
Effect	-1.42	-0.001	-0.207	-0.08	-0.054	-0.048	-0.004	-0.016
Mean	255	2955	108	338	302	471	886	5280
Elasticity	1.988	0.016	0.123	0.148	0.090	0.124	0.019	0.464

Source: GSS (1988/89/92/1998) and authors' calculations.

It appears that the parameters of the purely parametric index in Table 2 all have the expected negative sign, indicating that a prolonged growing season, higher prices and an increased farmsize all tend to lessen the need for indemnification. It also appears that the fit is fairly good, with an R^2 of 0.54. We report on R^2 for ease of reference but note that this is not the actual criterion used to derive optimal parameter values.

As regards the magnitude of the effects, the elasticity estimates (row 3) indicate that indemnification is more responsive to weather than to price shocks. For example, a 10 percent variation in the LGP resulting from say, a decrease by one month in the growing period of perennials leads to an increase of about 25 percent in the indemnification payment, corresponding to almost 50.000 cedi. With respect to prices, it appears that the index function is most responsive to the price of yam (elasticity of 0.14), followed by cassava and maize (around 0.12) and cocoyam, while the price of cocoa and tomatoes have less effect. This agrees with the relatively high vulnerability of staple crop producing farmers in the northern parts of Ghana, as compared to the better economic situation of cocoa and vegetable producing farmers in the southern parts.

Despite an adequate fit, the parametric index function has limited capacity to address poverty, as indicated in Table 3 below. It only reduces the simulated poverty incidence from 47 to 43 per cent. Nonetheless, we may note that, by virtue of its least absolute deviation property, it is likely to perform better than the more common synthetic specifications based on covariance matrices.

TABLE 3. Estimated Semi-Parametric index function, back-fitting procedure (β fixed)

n	R2	Poverty incidence (%)	Mean of α	Std dev of α	Mean of abs. errors	Regularization term
			Uninsured case			
	-	0.47	-	-	-	-
			Parametric index			
	0.54	0.43	-	-	107.2	-
			Semi-Parametric index $\lambda(n)$			
1	0.98	0.27	-2.5	7680	4.0	0
2	0.98	0.27	-2.6	6060	4.1	137
3	0.96	0.27	1.8	2769	8.7	501
4	0.79	0.28	10.4	703	30.8	1194
5	0.78	0.38	5.8	362	60.9	288

Source: GSS (1988/89/92/1998) and authors' calculations.

Table 3 shows that the semi-parametric specification achieves important improvements relative to the parametric form. The ability of the non-parametric part to adapt better the index-based insurance to the needs for indemnity payments comes to the fore in the second and third column. Without regularization, at $n = 1$, the fit is very good (R^2 of 0.98) but as expected, it gradually decreases with an increasing regularization, slowly approaching the fit of the purely parametric index function (R^2 of 0.54). By the same token, as compared to the parametric index and at moderate levels of regularization, the poverty incidence is substantially lowered to around 27 per cent (column 3).

We also remark that the index function estimated at $\lambda(1) = 0$ can be given a particular interpretation as it corresponds to the minimum level of farmers' basis risk of any index function based on the selected index variables and satisfying the self-financing constraint. In the prevailing case, this minimum is 24,000 cedi, comprising an assumed 20,000 induced by the η-margin plus an average absolute error of only about 4,000 from the ideal indemnification payment (column 6), as compared to a mean absolute error of 107,200 for the parametric index. An even lower η-margin would definitely reduce the value further, but eventually hit the limits of the spread in *x*-values.

The table also shows how, under regularization, the mean and standard deviation of the α-parameters decline significantly, reducing both the fit and the bumpiness. At the same time, the mean absolute error (first part of the objective) increases.

TABLE 4. Coefficients and elasticities of parametric part, for joint estimation of α and β

β	LGP	Cocoa	cassava	Yam	cocoyam	maize	tomatoes	Farm size
Effect	-1.52	-0.001	-0.192	-0.1	-0.075	-0.071	-0.006	-0.016
Elasticity	2.128	0.016	0.114	0.186	0.124	0.194	0.029	0.464

Source: GSS (1988/89/92/1998) and authors' calculations.

In Table 4, we report on α and β estimated jointly, focusing on the best regularization factor obtained at $\lambda(5)$ that reduces the variability of α_s to a standard deviation of 362, without causing a sharp decline in the fit. ($R^2 = 0.78$).

Comparing Table 4 with Table 2 indicates that β-parameters remain close to the values obtained from purely parametric estimation. In some cases, the interpretability of coefficients improves, as the elasticity with respect to the price of maize, very important for many poor farmers, and of the LGP becomes much stronger.

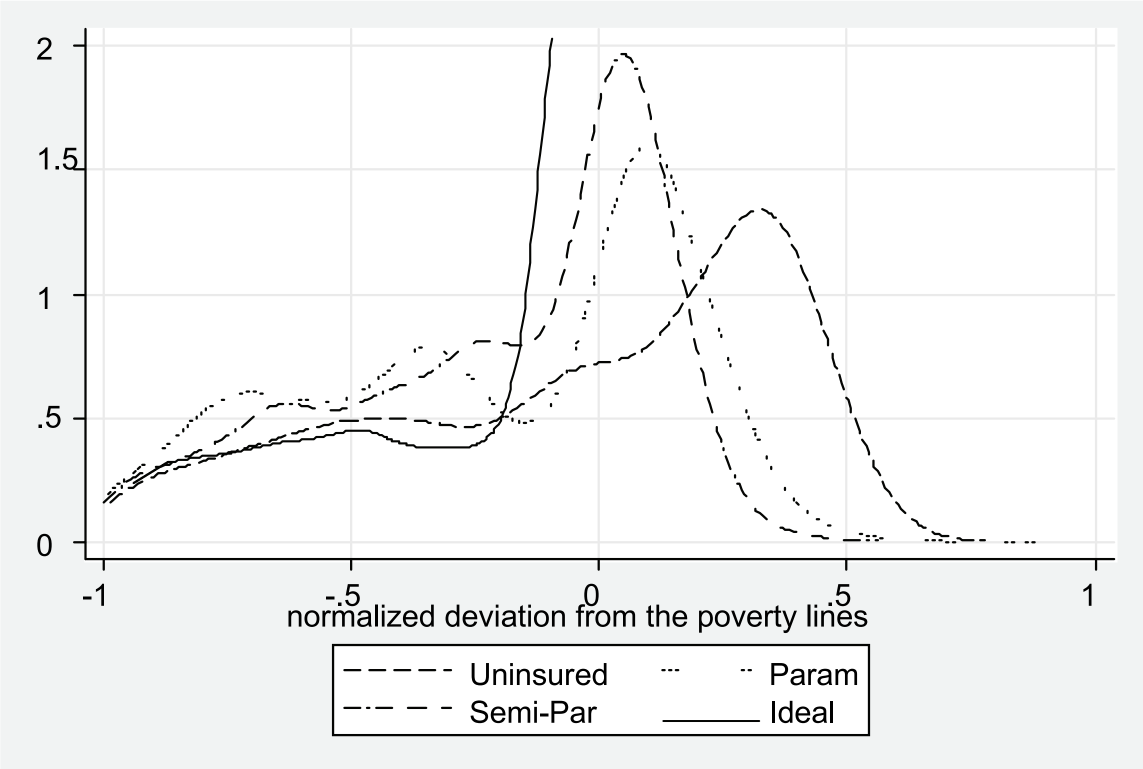
The similarity of the coefficients in both estimations suggests that the kernel terms of the support vectors of the non-parametric part are fairly orthogonal to the linear terms of the parametric part. This provides support for the interpretation of the kernel term as an additional explanatory element that offers a better model of the structural component of the error than is possible under normality assumptions, and more generally, under the assumptions of zero median and zero skewness.

As a further illustration of the performance under the various arrangements, Figure 2 compares deviations from the poverty line in the ideal insurance case (the continuous line), with the fully parametric index (the dotted line), the semi-parametric index function (the dashed-dotted line), and the uninsured case (the dashed line). Some interesting aspects come to the fore.

First, comparing the uninsured case with the ideal insurance, as expected, the probability of shortfall below the poverty line is reduced to zero in the ideal national scheme. However, due to premium payments, many farmers are moving towards the poverty line in many states of the world, and, for example, the share with an income exceeding twice the poverty line (1,750,000 cedi or 2 USD per capita per day) is only 13 percent, half of the percentage without insurance.

Second, comparing the uninsured case with the two index-based insurances, we see a tendency for shortfalls to diminish significantly but obviously much less than in the ideal case, where all shortfalls are eliminated: poverty prevalence decreases and the depth of poverty is reduced as well, as can be seen from the narrowing of the right-hand side tails. Since this narrowing is much more pronounced for the semi-parametric index, the semi-parametric form clearly outperforms the purely parametric one.

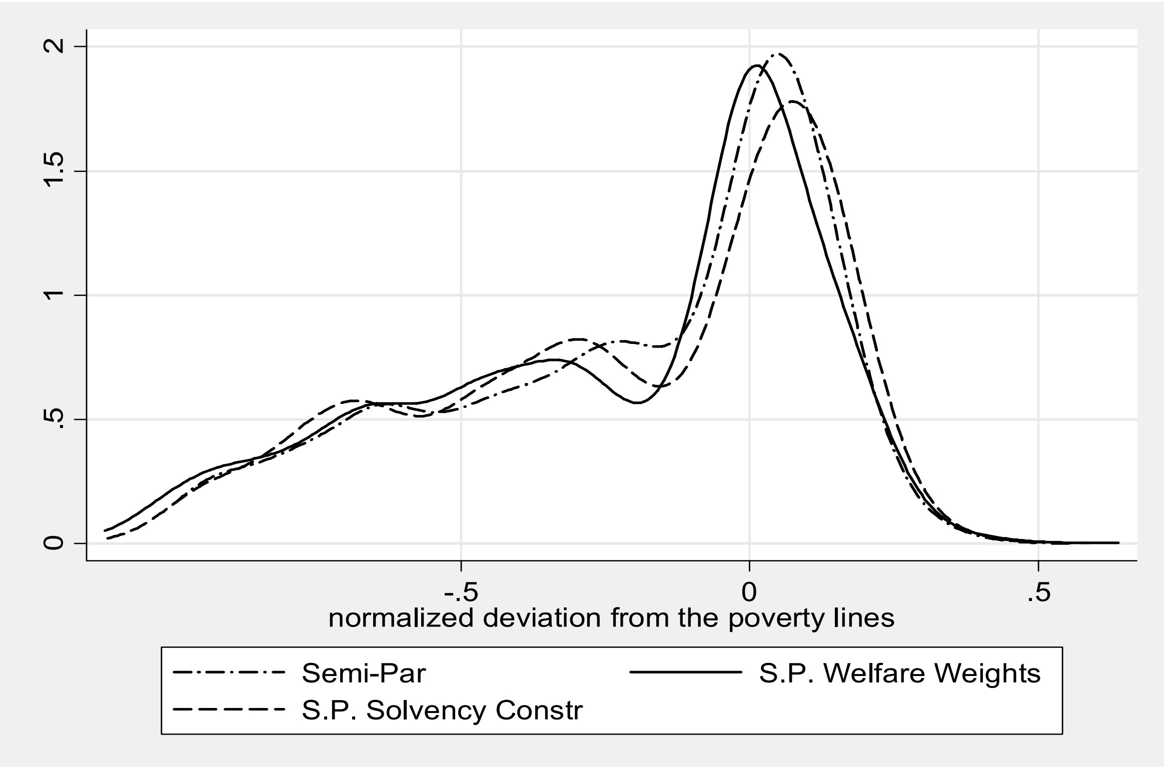
FIGURE 2. Parametric versus semi-parametric: Deviation from the poverty line (positive values=poverty): Uninsured income, index-based insurance (parametric/ semi-parametric) and ideal insurance



Source: GSS (1988/89/92/1998) and authors' calculations.

Next, in Figure 3, as an illustration of the flexibility of the modelling framework, and also to assess the stability of estimators in a structured way, we compare the distribution obtained from semi-parametric estimation with two variants, one (the continuous line) obtained by introducing some poverty-targeting weights in the objective of program (2.12), calculated as ratio of crop income over non-agricultural income divided by the farmsize, which biases the estimation of indemnification payments in favour of the poor, the other (the dashed line) by incorporating two solvency constraints within the program, so as to help avoiding liquidity squeezes that might arise after a succession of unfavourable years.

FIGURE 3. Solvency: Deviation from the poverty line (positive values=poverty): index-based semi-parametric insurance, index-based semi-parametric with weights on poverty and index-based semi-parametric with solvency constraints.

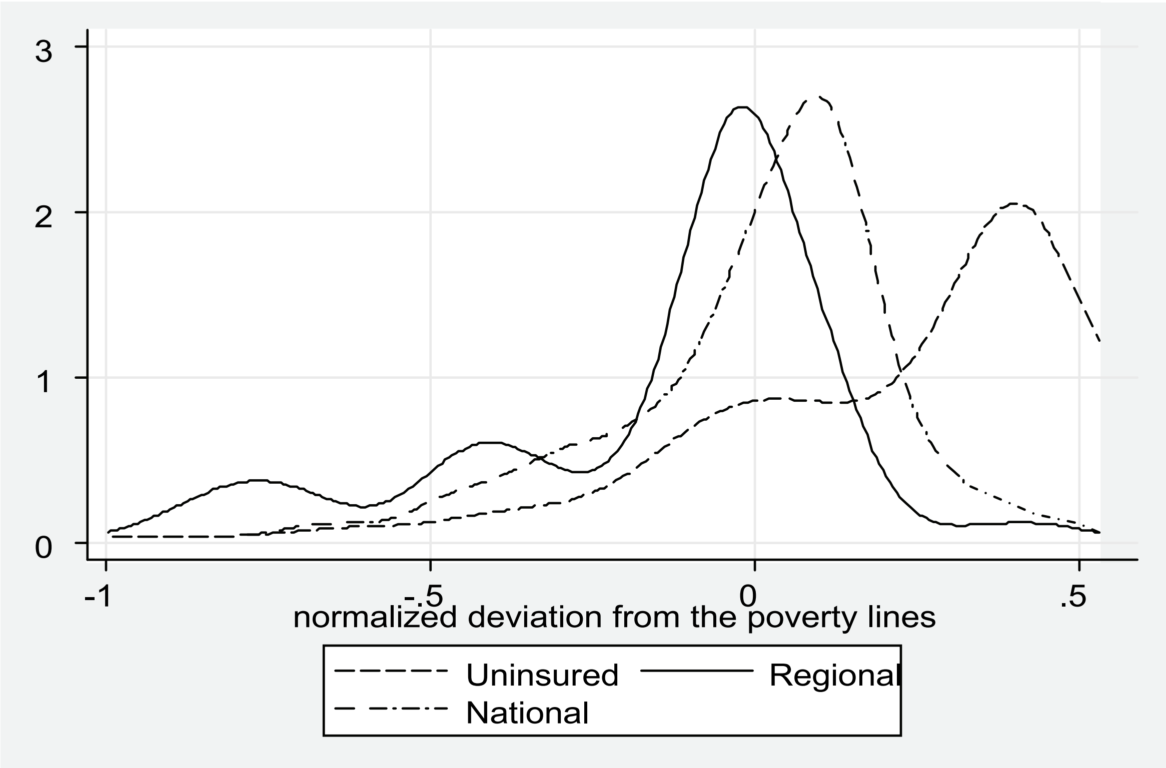


Source: GSS (1988/89/92/1998) and authors' calculations.

To allow for direct comparison, we estimate all three variants with the same regularization factor $\lambda(5)$. It appears that the coefficients are not much affected by this change, confirming their stability. Yet, the impact on poverty is noticeable. Introducing the bias towards the poor, yields a further reduction in the poverty rate from 38 to 34 percent, and significantly affected by this weighting are only those poor whose income was already shifted relatively close to the poverty lines before the bias was introduced. Hence, the weighting serves as targeting device.

In the other scenario, the introduction of solvency constraints operates as a restriction that reflects the limits of the insurer to transfer resources from good periods to bad ones. It raises the poverty rate to around 40 percent, about 2 percent above the base case.

Figure 4. Regional versus National: Deviation from the poverty line (positive values = poverty) in Upper East: index-based semi-parametric insurance (national-scheme), index-based semi-parametric insurance (regional scheme) and uninsured case



Source: GSS (1988/89/92/1998) and authors' calculations.

Lastly, Figure 4 compares the national scheme with a regional one for the Upper East, one of the poorest regions in the country, with a poverty rate of around 75 percent. Recall from Table 1 that in this region, coverage under the regional insurance scheme is only practicable with a significant external subsidy. Given this subsidy, the regional index-based insurance can offer a better adaptation to the local indemnification needs, reducing the regional poverty rate from 50 percent under the national arrangement to about 40 percent, with the subsidy contributing 35% of this reduction. This essentially indicates that LGP, prices and farmsize do not tell the whole story of vulnerability, and that further differentiation of insurance arrangements by region may be useful.

5. Summary and Conclusions

Our paper studies to what extent indexed-based insurance, price-weather insurance in particular, can contribute to poverty reduction, looking for ways to reduce the gap between both strands of literature.

Ongoing efforts to implement index-based insurance in developing countries indicate that participation rates are generally low and that self-financing is very modest. Leaving aside various explanations associated to institutional weaknesses such as corruption, we noted that this limited enthusiasm might also be due to the fact that the basis risk of the proposed insurance will tend to be relatively high as long as these remain based on functions that are formulated on a priori grounds, without much attention for the fit to the risk profiles of the target groups.

To address this aspect of the wider problem, we have proposed and implemented an approach for designing an index-based insurance that can optimally adapt to these profiles, while focusing on avoidance of catastrophic risk, defined as a fall below the poverty line. In addition, the contract ensures self-financing for a given level of subsidy and under assumed pooling of individual farmers within specified groups: quintiles, regions and the nation.

Next, we have considered a set of weather-price variables and constructed a function that comes as close as possible to the ideal individualized insurance by minimizing the basis risk associated to index-based insurance. For this purpose, we adapted a technique from SV-regression while allowing for financing restrictions that inherits the flexibility and the strong convergence properties of SV-regression. In addition, the absolute value criterion of our formulation enabled us to optimize in a fully integrated way from the perspective of the risk-minimizing farmer, as opposed to fitting a curve from the predictor’s perspective, hence avoiding the traditional dichotomy between formulation of a micro-model of an income risk minimizing farmers on the one hand and econometric estimation with a likelihood risk on the other. The numerical application has been conducted in a dedicated software package that can flexibly accommodate bagging as well as various extensions of the constraints beyond self-financing.

At the practical level, we ran the estimation on a sample of 100 representative Ghanaian households over 26 possible states of the world drawn from historically observed conditions. The parametric part of the SV-regression showed the expected signs, with indemnification requirements being most responsive to low rainfall and low staple prices. This might reflect the relatively high vulnerability of staple crop farmers in the Northern part of Ghana as compared to the better-off situation of cash-crop farmers in the Southern part.

Our comparison of parametric with semi-parametric index functions showed that the non-parametric part improves the capacity to adapt to the required indemnity payments. We also studied how, under increased regularization, the fit of the semi-parametric index approaches the purely parametric one, reducing the fit but also the bumpiness of the prediction. The simulated effects on poverty were substantial. Poverty prevalence was reduced on average but the depth of poverty decreased as well, especially for the semi-parametric index at a moderate level of regularization.

Finally, regarding work in progress, we intend to assess the robustness further by applying bagging through estimation on series of sub-samples.

REFERENCES

Barnett, B.J., C.B. Barrett, and J.R. Skees (2006) "Poverty traps and index-based risk transfer products". Mimeo.

Beaudry, P., and D.A. Green (2000) "Cohort patterns in Canadian earnings: assessing the role of skill premia in inequality trends". *Canadian Journal of Economics* 33(4): 907-36

Chapagain, A.Y., and A.K. Hoekstra (2004) *Water Footprints of Nations*. UN/WWAP, UN World Water Development Report. New York: United Nations

Deaton, A.S. (1985) "Panel data from time series of cross-sections". *Journal of Econometrics* 30(1):109-26.

Decaluve, B., L. Savard, and E. Thorbecke (2001) "General Equilibrium approach for poverty analysis" Mimeo.

Dercon, S. (1996) "Risk, crop choice, and savings: Evidence from Tanzania". *Economic Development and Cultural Change* 44, 485-513.

Dercon, S. ed. (2005) *Insurance Against Poverty*. Oxford: Oxford University Press.

Elbers, C., J.W. Gunning, and B. Kinsey (2007) "Growth and risk: Methodology and micro evidence". *World Bank Economic Review* 21(1): 1-20.

Ermoliev, Y.M., T.Y. Ermolieva, G.J. MacDonald, and V.I. Norkin (2000) "Stochastic optimization of insurance portfolios for managing exposure to catastrophic risks". *Annals of Operations Research* 99: 207-25.

Fafchamps, M. (1992) "Cash crop production, food price volatility, and the rural market integration in the Third World". *American Journal of Agricultural Economics* 74: 90-9.

Food and Agriculture Organisation (2005) "Fertilizer use by crop in Ghana" Rome: Food and Agriculture Organization of the United Nations.

Fofack H. (2000) "Combining light monitoring surveys with integrated surveys to improve targeting for poverty reduction: The case of Ghana" *World Bank Economic Review* 14(3): 195-219.

Gilonia, A., J.S. Simonoff and B. Senguptac (2006) "Robust weighted LAD regression". *Computational Statistics and Data Analysis* 50: 3124-40.

GMI (2006) *Monthly Rainfall in 72 Synoptic Rainfall Stations*. Accra: Ghana Meteorological Institute.

GSS (1989/1995/2000) *Reports on Ghana Living Standards Survey (Round 1 to 4)*. Accra: Ghana Statistical Service.

GSS (1987/2001) *Reports on Ghana Population Census (1984 and 2000)*. Accra: Ghana Statistical Service.

GSS (2005) *Quarterly Digest of Statistics*. Accra: Ghana Statistical Service.

Goodwin, B. K. and O. Mahul (2004) "Risk modelling concepts relating to the design and rating of agricultural insurance contracts". Policy Research Working Paper 3392 Washington D.C.: World Bank.

Haerdle, W. (1995) "Applied Nonparametric Regression" Cambridge UK: Cambridge University Press.

Herbrich, R. (2002) *Learning kernel classifiers. Theory and Algorithms*. Cambridge Ma: MIT Press.

Hess, U., and J. Syroka (2005) "Weather-based insurance in Southern Africa: the Malawi case study". World Bank Agricultural and Rural Development Discussion Paper 13, Washington DC: World Bank.

Huber, P.J. (1981) *Robust Statistics*. New York: Wiley.

Kalavkonda V. and O. Mahul (2005) "Crop insurance in Karnataka". World Bank Policy Research Working Paper 3654. Washington D.C.: World Bank.

Keyzer M.A. (2005) "Rule based and support vector regression and classification algorithms for joint processing of census, map survey and district data". SOW-VU Working Paper 05-01. Centre for World Food Studies, Vrije Universiteit Amsterdam.

Kimeldorf, G.S. and G. Wahba (1970) "A correspondence between Bayesian estimation on stochastic processes and smoothing by splines". *Annals of Mathematical Statistics* 45: 495-502.

Koenker, R., and Bassett, G. W. (1978) "Regression quantiles". *Econometrica* 46: 33–50.

Mahul O. and B.D. Wright (2003) "Designing optimal crop revenue insurance". *American Journal of Agricultural Economics* 85(3): 580-9.

Norkin and Keyzer (2007) "On convergence and stability of kernel learning quantile estimators", Glushkov Institute of Cybernetics, Kiev, and Centre for World Food Studies, Amsterdam. Mimeo.

Sarris A. (2002) "The demand for commodity insurance by developing country agricultural producers: theory and an application to cocoa in Ghana" Policy Research Working Paper 2887 Washington D.C.: World Bank.

Schoelkopf, B. and A.J. Smola (2002) *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge Ma: MIT Press.

Skees, J.R, S. Gober, P. Varangis, R. Lester and V. Kalavakonda (2001) "Developing rainfall-based index insurance in Morocco". Policy Research Working Paper 2577. Washington D.C.: World Bank.

Skees, J.R., P. Hazell and M. Miranda (1999) "New approaches to crop yield insurance in developing countries". EPDT Discussion Paper 55. Washington D.C.: International Food Policy Research Institute.

Skees, J., P. Varangis, D.F. Larson and P. Siegel (2005) "Can financial markets be tapped to help poor people cope with weather risks?" In: S. Dercon ed. *Insurance Against Poverty*.

Takeuchi I., Q.V. Le, T. Sears, and A.J. Smola (2005) "Nonparametric quantile estimation", *Journal of Machine Learning Research* 7: 1001-33.

Teal F. and M. Vigneri (2004) "Production changes in Ghana cocoa farming households under market reforms" Mimeo.

Vapnik, V.N. (1998) *Statistical Learning Theory*. New York: Wiley.

Veeramani, V.N., L.J. Maynard and J.R. Skees (2005) "Assessment of the risk management potential of a rainfall based insurance index and rainfall options in Andhra Pradesh, India". *Indian Journal of Economics & Business* 4(1): 195-208.

Zant W. (2005) "Is index insurance useful for cash crop growers? Simulating production and price index insurance for Indian smallholder pepper growers". Mimeo.

APPENDIX to Section III

FARM HOUSEHOLD CHARACTERISTICS

TABLE 1. Summary statistics on farm household characteristics in Ghana, by region

Region	Income (1000 cedi per capita)	Crop income (1000 cedi per capita)	Average Shortfall below the poverty line (1000 cedi per capita)	Farm size (ha per capita)	Farm population (%)	Farmed land (%)	LGP annual crops (days)	LGP perennial crops (days)
Ashanti	1439	865	57	0.49	8.4	18.4	217	276
Brong Ahafo	1187	1049	99	0.72	10.3	11.6	219	261
Central	1262	940	79	0.61	9.3	9.8	165	255
Eastern	1053	681	91	0.39	8.6	14.1	226	282
Greater Accra	1500	581	49	0.34	6.3	2.3	112	148
Northern	717	398	193	0.54	13.1	10.3	219	222
Upper East	698	375	223	0.47	11.2	8.2	177	183
Upper West	680	498	222	0.61	13.7	4.1	180	196
Volta	949	503	146	0.33	8.2	10.4	232	264
Western	1629	1375	24	0.70	10.9	10.7	212	306
Ghana	1439	865	57	0.49	8.4	18.4	217	276

Source: GSS (1989, 1995, 2000), FAO (1979) and authors calculations. Income figures are deflated to March 1999 Accra prices (GSS, 2000), the exchange at the time was 2394 cedi/$

Section 3 describes the steps of data base compilation. Here Table 1 presents summary statistics averaged by regions. The distribution of total income and crop income, (column 1 and 2) strikingly highlights the sharp economic North-South divide in the country. Farmers in the three Northern regions (Northern, Upper East and Upper West) have an average income below the 700.000 cedi poverty line (column 1 and 5), while in the rest of country, the average income is above (Volta) or significantly exceeds this threshold. We also remark that on average crop income accounts for three quarters of total income of farm households. Spatially this share ranges between less than 50% in Greater Accra, to more than two thirds in the rest of the country, rising as one moves into the Western part of the country.

Column 3 of the table indicates that on average shortfall below the poverty line occurs for some farm groups in all regions in some of the simulated states of the world, except in the Greater Accra region, where farm households are relatively rich and off-farm income plays a major role.

Finally, columns 7 and 8 point to significant differences also in the rainfall pattern and consequently in the Length of Growing Period, LGP[3], as is illustrated further in Figure 2. This is reflected in the variation of cropping pattern across the regions. In the North, Upper East and West, due to shorter and less dependable precipitation, farmers typically cultivate annual crops and choose varieties that need less water, such as sorghum, millet, beans and nuts. In the Southern half of the country where precipitation is more frequent, they turn to perennial crops such as cocoa, palm oil, yam, cocoyam and plantains.

[3] LGP is the period during which precipitation exceeding half of the reference evapo transpiration. Under rainfed conditions this is considered sufficient to meet the water requirements of a crop (FAO, 1979).

APPENDIX to section IV: COMPUTATIONAL ASPECTS

In many practical applications, including ours, the dimensionality of the program becomes an issue, as the (square) matrix with kernel elements becomes large. There are various ways out. One is to opt for a more powerful software tool, which does not seem advisable in our case since we seek to present a tool that could run on laptops in Africa. The second is to choose a kernel that more strongly penalizes dissimilarity (has a small window size) and hence has more zeroes in its kernel matrix but this discards interdependencies in the data. A third option is to limit the number of representative households, while adjusting their poverty threshold so as to respect the poverty measurement in the underlying data set. This is the approach followed as discussed in Section IV.

In addition, we also mention bagging as a fourth option. Bagging amounts to creation, by drawing (with or without replacement) from the data set, a series of random samples $S_o \subset \{1,.....,S\}$ from the full data set, while keeping the financing constraint, since it is a single equation that can, because of the factorization, be maintained for the full set. This leads to the program:

$$\min_{\xi_s,\eta\geq 0,\alpha_s,\beta_j} \quad \sum_{s\in S_o} n^o_{i_s}\xi_s + \frac{\lambda_{S_o}}{2}\sum_{s\in S_o}\sum_{r\in S_o}\alpha_s\alpha_r k(x_s,x_r) + \vartheta\eta$$

subject to

$$y_s \leq \sum_r \alpha_r k(x_r,x_s) + \sum_j \beta_j\phi_j(x_s) + \xi_s + \eta, \qquad s\in S_o \qquad \text{(A.1)}$$

$$y_s \geq \sum_r \alpha_r k(x_r,x_s) + \sum_j \beta_j\phi_j(x_s) - \xi_s - \eta, \qquad s\in S_o$$

$$\sum_{r\in S_o}\alpha_r\Big[\sum_{s\in S}\frac{n_{i_s}}{L}k(x_r,x_s)\Big] + \sum_j \beta_j\Big[\sum_{s\in S}\frac{n_{i_s}}{L}\phi_j(x_s)\Big] = \bar{\sigma}$$

where $n^o_{i_s}$ is the given weight on observation s, and terms in square brackets are evaluated beforehand. This series generates a distribution of y-predictions and β-estimates, whose moments can be computed to test for stability and out-of-sample robustness, as is commonly done in SV-regression. Indeed, (3.14) is a natural complement of the estimation on the full sample.

Consequently, in our case study for Ghana we keep a sample of size $S = 2600$, which corresponds to around 3.4 million quadratic terms, of which around 1 million are non-zero by appropriate choice of window size in the kernel matrix. After that, we intend to establish the robustness of our estimates by expanding the constraint set with solvency constraints and by a series of sub-samples in program (3.14) implemented with a broader window size that explores more interdependencies, but this is work in progress.

Properties

Finally, we briefly discuss to which extent the resulting estimator inherits the convergence properties of SV-regression estimators. In section II the discussion refers to the full sample in equation (2.12) but applies to (A.1) as well, and considers a sequence of ever increasing full samples. It can be seen from (2.12) that even though the contract estimated by the function is not individualized, the program seeks to fit it fully to all observations at individual level. One question is then, whether it can achieve this, say, in the absence of soft margin, and when regularization vanishes, and the other, whether the sequence of functions f^{S_ℓ} obtained by (2.12) for sets $S_\ell \subset \{1,.....,S\}$ of iid observations indexed ℓ, will for the size of the underlying data set

$L \to \infty$, regularization factor $\lambda_{S_\ell} \to 0$ and soft margin penalty $\vartheta \to \infty$ converge (in specific sense) to the true function *f*.

Regarding the first question, semi-parametric form (see Section II eq. 2.11) owes its popularity to the capacity to fit given data, essentially because the matrix with elements $[k(x_s,x_r)]$, $r,s \in S_o$, which is known as the Gram matrix, is for any kernel function known to be positive semi-definite symmetric by definition, and to offer a measure proximity between two observations x_s and x_r. Specifically, while extreme proximity with identical *x*-values for all observations will reduce the rank of this Gram matrix to unity, extreme distance, with only the diagonal terms as non-zero values, leads to a diagonal form, and hence to full rank. Full rank in turn means that the function has full capacity to fit the given data set: $y = K\alpha + \Phi\beta$ can for given can for given y and $\Phi\beta$ always find an α-value that solves it. Yet, in our case of an index function with common anchors, the rank will often be deficient, because there are several households sharing the same *x*-values. In an econometric application, this might be seen as a limitation of the data set. Here this incapacity to match every individual's needs is an inherent limitation of the construct of index-based insurance with non-individualized contracts.

Reference

Food and Agriculture Organisation (FAO) 1979. "Yield response to water. Drainage and Irrigation" Paper 33. Rome: Food and Agriculture Organization of the United Nations.

The Centre for World Food Studies (Dutch acronym SOW-VU) is a research institute related to the Department of Economics and Econometrics of the Vrije Universiteit Amsterdam. It was established in 1977 and engages in quantitative analyses to support national and international policy formulation in the areas of food, agriculture and development cooperation.

SOW-VU's research is directed towards the theoretical and empirical assessment of the mechanisms which determine food production, food consumption and nutritional status. Its main activities concern the design and application of regional and national models which put special emphasis on the food and agricultural sector. An analysis of the behaviour and options of socio-economic groups, including their response to price and investment policies and to externally induced changes, can contribute to the evaluation of alternative development strategies.

SOW-VU emphasizes the need to collaborate with local researchers and policy makers and to increase their planning capacity.

SOW-VU's research record consists of a series of staff working papers (for mainly internal use), research memoranda (refereed) and research reports (refereed, prepared through team work).

Centre for World Food Studies
SOW-VU
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

Telephone (31) 20 – 598 9321
Telefax (31) 20 – 598 9325
Email pm@sow.vu.nl
www http://www.sow.vu.nl/