Supply Chain Analysis of Contract Farming (Unabridged Version)

Awi Federgruen
Graduate School of Business, Columbia University, New York, New York 10027, af7@columbia.edu

Upmanu Lall
Columbia Water Center and The Department of Earth and Environmental Engineering, Columbia University, New York, New York 10027, ula2@columbia.edu

A. Serdar Şimşek
Naveen Jindal School of Management, The University of Texas at Dallas, Richardson, TX 75080, serdar.simsek@utdallas.edu

Contract farming is a growing practice in developing countries and first-world economies, alike. It generates necessary guarantees to sustain the continued operations of vulnerable farmers while enabling the manufacturers to manage the aggregate supply and price risk. We consider a single manufacturer who owns several manufacturing plants, each with a random demand for the crop. The manufacturer selects a set of farmers to offer a menu of contracts, which is exogenously specified or endogenously determined. Each "selected" farmer chooses a contract from this menu in advance of the growing season. After the growing season, under known demands and supplies, the manufacturer minimizes the distribution costs from the selected farmers to the production facilities. We formulate this problem as a Stackelberg game with asymmetric information, where the manufacturer is the leader and the farmers are followers. The manufacturer's problem is a two-stage stochastic planning program for which we develop two solution approaches. We have applied our model to problem instances anchored on data from a large manufacturer of potato chips contracting with thousands of small farmers in India. We report on the performance of the solution methods compared to a lower bound based on the Lagrangean dual of the problem and show that the optimality gap is below 1%, for problem instances with 1,000 potential farmers. We also show how our model can be used to gain various managerial insights. As an example, when constructing the contract menu endogenously, often a small number of contract options suffices, depending on the degree of heterogeneity among the farmer pool. Thus, relatively simple menus often suffice.

Key words: contract farming; food and water security; optimization under uncertainty; stochastic model applications; agriculture/food industry

---

1. Introduction and Summary

Contract farming is an emerging practice in developing countries such as India, Brazil, and Turkey. Food processing companies, hereafter referred to as manufacturers, contract with a large number of relatively small and financially challenged farmers. The industry distinguishes between two types of contracts: under production contracts, the buyer owns the product while it is being produced and the farmer is compensated for the services he provides; in a marketing contract, the farmer maintains ownership over the crops but, in its simplest form, the manufacturer guarantees to buy (up to) a given quantity of the farmer's crop in a season, for a set guaranteed price. More specifically, the manufacturer buys either the farmer's realized yield or the specified contract volume, whichever is lower. In this paper, we confine ourselves to marketing contracts.

A manufacturer often contracts with hundreds or thousands of farmers. These farmers are dispersed across different regions, often with vastly different climate and soil conditions, water supplies and farm sizes.

Contract farming protects the farmer against the many risks he is facing, in particular the risk of volatile crop prices on the commodity spot markets, yield risks and the difficulties of finding buyers for their crops. Indeed, commodity prices are very volatile; for example, in 2013, the spot price for potatoes more than doubled compared to its value a year earlier; see Table 6 in Appendix F for another example. Yield risks arise due to factors such as uncertain weather conditions and high volatility in the water supply.

However, contract farming is also highly desirable, if not outright critical, for the manufacturer, and there are several reasons for this. First, contracts are essential to enable or incentivize farmers to remain in the market, given the daunting risks they experience otherwise. Second, the high volatility of spot markets presents a major problem for the manufacturer as well. Third, while airlines, for example, are able to hedge against the risk of fluctuating fuel prices by investing in call options, such hedging instruments are far less prevalent for agricultural commodities, where the option markets have limited liquidity. In India, for example, option contracts on commodities are not yet permitted, even though the Security Exchange Board of India has considered their introduction (see Rajib 2014, 2015).

Fourth, and perhaps most importantly, manufacturers typically need their input commodity (e.g. potatoes) to comply with a series of detailed specifications. These specifications, for example the use of specific seeds, can be built into a contract, but, on the spot market, it is very hard to find (sufficient) supplies that conform with them (see MacDonald et al. 2004, p. 29). The contracts also enable the buyer to provide equipment, raw materials, and financing to the typically cash-strapped farmers. Finally, quality inspections are hard to perform in the context of spot market transactions, but are naturally built into the farm contracts.

Take, as an example, the Starbucks company. Its most important input is unroasted coffee beans. In 2014, Starbucks bought more than half a billion pounds of these beans, contracting with more than 300,000 farmers worldwide in Latin America, Africa, and Asia; see Gruley and Patton (2014). In its contractual arrangements, Starbucks often specifies bean varieties, including specific new hybrid seeds. It also trains the farmers it contracts with, to standardize their methods in efficient, sustainable ways under the company's Coffee And Farmer Equity (C.A.F.E.) Practice program.

Monitoring and enforcing compliance is a challenge in all of contract farming, but no more than in other industrial contracts with upfront price or quantity commitments. If anything, monitoring of yields in open fields is easier than in industrial settings where products are manufactured behind closed doors; it is often done using satellite remote sensing or drones. Much thought has been given to best practices in this area. As an example, a "Legal Guide on Contract Farming" was developed, in 2013 – 2015, by the International Institute for the Unification of Private Law (UNIDROIT), in partnership with the FAO. The FAO also put out a document "Contract farming – partnerships for growth"; see Chapter 6, "Monitoring performance", for detailed and effective monitoring procedures. Thus, contract compliance can be and is monitored effectively. Beyond the formal justice system, this creates a natural deterrence for a farmer to violate the agreement, thereby jeopardizing or foregoing future contracts in upcoming seasons.

Thus, contract farming is a novel supply chain arrangement that benefits all parties. In particular, it generates necessary guarantees to sustain the continued operations of very vulnerable suppliers (farmers) – a major socio-economic goal for developing countries.

Moreover, contract farming is a common and growing trend in *first-world* countries as well. It covered no less than 39% of US agricultural production in 2008, up from 11% in 1969 and 28% in 1991; no less than 90% of sugar beet and tobacco production was conducted via contract farming; see MacDonald and Korb (2011) and Table 9 in Appendix G. As in emerging countries, American farmers are exposed to production risks resulting from droughts, frost, hail, diseases, and insect infestations, among other unusual events, and the above severe volatility in commodity prices on the spot market, providing a major incentive for contract farming.

However, the devil is in the tactical details. To ensure an adequate supply of the crops, how many farmers should the manufacturer contract with? How should specific farmers be selected? What is the impact of the offered menu of price-quantity contracts and how should the menu be chosen? If the manufacturer has a network of production facilities, which farmers should supply which of these facilities?

To answer these interdependent questions, we analyze the following model and develop effective solution methods: There is a single manufacturer who owns several manufacturing plants, each with a specific, albeit random demand for the crop. The manufacturer incurs two types of cost: procurement and distribution costs. She offers a specific menu of contracts to a selected set of farmers in advance of the growing season. A contract specifies the unit price and quantity (among other terms and specifications). The actual purchase amount is the minimum of the committed quantity and the realized yield. Offering a *menu* of contracts is essential under *information asymmetry*, i.e., when the manufacturer has less than perfect knowledge of all of the farmers' characteristics, for example their yield characteristics.

Each "selected" farmer chooses a contract from this menu which results in the highest expected profits for him, among those that satisfy certain participation or risk aversion constraints. Our model assumes that the water supply is the principal random factor affecting a farmer's total yield, primarily due to uncertainty about the rainfall process. In some applications, farmers have access to surface water (from sources such as lakes and rivers) as a second source of water supply. In addition, some farmers are able to supplement the water supply by drawing ground water from a capacitated well, incurring additional energy and water costs. As with the rainfall and surface water quantity, the well's capacity in any given season is, often, random as well. The yield of the fields is a function of the total water supply, i.e., the aggregate of the rainfall, surface and drawn ground water. With minor adjustments, the model can be applied to other –or multiple– yield risk factors.

During the season, the various random factors become known: first, the rainfall, surface water, and the well capacity of each of the farmers become revealed. Thereafter, each contracted farmer determines how much water to draw from a capacitated well (if available). This determines the farmer's total yield, a (non-linear) deterministic function of the farmer's combined water supply. The manufacturer observes, by the end of the season, its product demand at each plant and the realized supplies of all contracted farmers. She then determines an optimal distribution plan to supply its plants from the contracted farmers, drawing from the spot market or an external source in case of a nationwide shortage of farmers' supplies.

In selecting how many and what specific collection of farmers to contract with, the manufacturer wishes to minimize expected aggregate costs subject to a *coverage constraint* ensuring, with a given high likelihood, that the nationwide aggregate supply matches or exceeds aggregate nationwide demand. The importance of this constraint follows from the fact that, for all the reasons outlined above, the manufacturer cannot rely on the spot market as a supply source other than as a rare emergency backup solution, at worst to be used for a small part of the overall supply.

We formulate this problem as a Stackelberg game, where the manufacturer is the leader and the farmers are followers. The manufacturer has imperfect knowledge of part of the parameters in the distributions that describe the farmers. In other words, some of the farmers' information is *private*, giving rise to a game with *asymmetric information.*

The farmers only need to know the parameters pertaining to their *own* farm and, at first, select a contract from the menu. Various NGO's and government agencies assist farmers with their contract decisions. An example is SNV, working together with the Food and Agriculture Organization (FAO) of the United Nations; see e.g., Sango et al. (2016) reporting on their contract farming support program in Zimbabwe.

In our base model, the menu of contracts is exogenously specified. (A different menu may be offered in different regions and to farmers of different land sizes.) However, we also show how the manufacturer may choose the menu of contracts, *endogenously*, along with her other decision elements, so as to minimize her expected costs.

Even for a given menu of contracts, the manufacturer's problem resulting from the farmers' best responses to the offered menu is a complex two-stage stochastic program: she first has to select the set of farmers to work with in the upcoming growing season, and after the growing season, under known demands and supplies, she determines the distribution plan from the selected farmers to the production facilities; this to minimize the aggregate of the expected procurement and distribution costs. An important feature of this two-stage stochastic program is the aforementioned coverage constraint. As mentioned, we also show how an optimal menu of contracts may be selected along with the manufacturer's remaining decision elements. In doing so, we follow a standard paradigm in contract theory, see e.g., Bolton and Dewatripont (2005), Lovejoy (2006) or Lutze and Ozer (2008).

Under a given menu of contracts, the manufacturer's two-stage stochastic planning problem – incorporating the farmers' best response functions– may be formulated as that of minimizing a supermodular set function subject to the above coverage constraint as a single side constraint. In other words, we view the outcome of the stochastic program as a function of the set of selected farmers. A set function is supermodular if the cost savings resulting from the addition of a given farmer decrease when this farmer is added to a larger collection of farmers. We develop two algorithmic approaches to solve the manufacturer's problem, as well as a lower bound for the optimal cost value, based on the Lagrangean dual resulting from the Lagrangean relaxation of a pair of quadratic constraints that are equivalent to the above probabilistic coverage constraint.

We have applied our model to problem instances that are calibrated to data from a large manufacturer of potato chips contracting with thousands of small potato farmers in India.

In summary, this paper makes the following contributions:

(1) To our knowledge, ours is the first paper to propose a farmer selection model to be used by a manufacturer contracting with hundreds or thousands of farmers, to be chosen from an even larger set of potential suppliers. The model assumes that each of the selected farmers is offered a menu of contracts, which may be exogenously specified or endogenously determined along with the selection of the farmers. The model is analyzed as a Stackelberg game with asymmetric information, with the manufacturer as the leader and the selected farmers as the followers, each selecting the contract which maximizes his expected profit subject to specific risk constraints. The manufacturer's selection problem incorporates the end-of-the-season problem of optimal distribution from the selected farmers to the plants as a two-stage stochastic program, since the farmers' locations represent an important criterion for this selection.

(2) We develop two approximate algorithmic approaches to solve realistic, large-scale problem instances in a very modest amount of time: instances with a thousand potential farmers can be solved in a few minutes of CPU time, when implemented in a standard laptop platform. The average

value of (an upper bound for) the optimality gap is less than 1%. This means that the heuristics are indeed close to optimal and the lower bounds are close to exact. Moreover, the optimality gap, always less than 1.5% for the best of the two algorithms, declines as the problem size increases.

(3) Our numerical studies have identified several interesting qualitative insights, the most important of which are summarized in the Conclusions section.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature. Section 3 develops our model. Section 4 characterizes the solution to the farmers' best response problem in response to the offered menu of contracts. Section 5 shows how the menu(s) of contracts may be endogenously determined. Section 6 develops the two algorithmic approaches for the manufacturer's optimization problem. Section 7 reports on a large scale numerical study. Our major conclusions are summarized in Section 8.

2. Literature Review

This paper contributes to the recent literature on supplier selection and diversification. Almost all of this literature has been motivated by industrial or public health applications. The initial literature on procurement problems under random yields confined itself to models with a single, albeit unreliable supplier. See Grosfeld-Nir and Gerchak (2004) for a survey of the literature up to 2004. Several papers, in the nineties and the new millennium, addressed procurement strategies when there are *two* or an arbitrary number of potential suppliers, respectively. See Federgruen and Yang (2011) for a recent survey on this literature steam.

There are a few papers with endogenously determined yield distributions, in particular: Liu et al. (2010) consider a newsvendor model and a single supplier with a stochastically proportional yield; the authors investigate the impact of improving the yield factor distribution and replacing it by a stochastically superior one. Federgruen and Yang (2009) address yield factor distributions that are endogenously determined by any number of suppliers as part of a competition game. Wang et al. (2010) assume, within a two-supplier random capacity model, that the suppliers' capacity loss distributions may be controlled by selecting an effort level from a continuous spectrum. In a two-stage process, the purchasing firm first decides how much effort to invest to improve the capacity loss distribution of both suppliers; after observing whether these efforts have succeeded, the purchasing firm decides on order quantities to cover a single season demand volume.

Our model differentiates itself from the above existing literature, by addressing several important complications and generalizations: (i) each farmer/supplier is offered a menu of possible contracts rather than a single order quantity at a single per unit cost price; (ii) after selecting a target delivery quantity from the menu, the farmer can, at least partially, control what fraction thereof is generated and sold to the buyer, by selecting the water supply from accessible wells to complement his random rainfall (and surface water, if applicable); (iii) rather than servicing a single location with random demand, we consider settings where random demands occur at an arbitrary number of locations, and the purchased supply quantities need to be distributed to these demand points so as to minimize aggregate distribution costs. Another important distinction in the contract farming context is that the number of potential (and actually retained) suppliers is in the hundreds or thousands, while in the above industrial applications, the number of potential suppliers is typically below 10. The qualitative implication of this distinction is that, in our context, supplier selection cannot be undertaken by a brute force enumeration of all possible sets of suppliers. At the same time, our model addresses a single season, while several of the above papers deal with models with multiple procurement opportunities and inventories carried from one period to the next.

Several papers have addressed procurement strategies for food processors. Devalkar et al. (2011) consider a multi-period setting in which the manufacturer procures its input commodity entirely from the spot market where it is subject to a spot price process which is stochastic but of a known distributional form. The authors were motivated by ITC, one of India's largest private sector food processing companies, and its procurement of soybeans for the production of soybean oil and meal. Since all inputs are purchased on the spot market, the Devalkar et al. (2011) model does not involve the selection of supply sources and associated contracts and the paper does not need to contend with yield uncertainty and distribution costs, as in our model addressing contract farming. Similarly, Tan and Çömden (2012) present a planning methodology for a manufacturer wanting to match the random supply of premium fruits and vegetables with random retailer demand for a *given* set of contracted farms, operating under a *given* price contract.

Similar to our paper, Chaturvedi and de Albeniz (2011) consider a supplier selection problem for a manufacturer, where the suppliers face uncertain yields and there is information asymmetry between the manufacturer and the suppliers. However, the yield uncertainty is assumed to be of a binary nature: either the supplier succeeds in delivering the full order size, or he delivers nothing. This representation of the yield factor does not fit the farmers' yield uncertainty, where it varies *continuously* and depends, in an intricate manner, on the rainfall and other sources of water supply. Rather than ensuring, with a coverage chance constraint, that the total supply covers a total random demand volume, the authors assume that the manufacturer wishes to maximize the expected value of a concave increasing function of the *realized* supply. Moreover, in Chaturvedi and de Albeniz (2011) only the aggregate realized supply matters, while in our model, given the importance of distribution costs, its geographic dispersion across the country is of essential importance as well.

Boyabatli et al. (2011) assume that the manufacturer may procure her input commodity either from upfront contracts or from the end-of-the-season spot market; both sources are viewed as equally desirable except for possible price differentials. The demand volume is deterministically given. The (uniform) unit contract price is linearly dependent on the end-of-the-season spot price, however with an upper and lower bound cap. (This is apparently a common practice in the beef industry.) The quantity bought in an upfront contract is assumed to become available without any yield uncertainty.

Huh and Lall (2013) endogenize the stochastic yield rate by incorporating irrigation decisions under rainfall uncertainty, similar to our treatment of the farmers' yield. Their model considers a single farmer producing for a single market, who can allocate his land among several possible crops, with uncertain commodity prices. A forward contract is offered to the farmer for a subset of the possible crops. Similarly, Huh et al. (2012) address contract farming by a single location manufacturer who has access to a pool of identical local farmers, all experiencing the exact same rainfall per acre; as opposed to a single farmer. de Zegher et al. (2017) present a contract farming model in the wool industry, with a *given* set of contracted farmers. Each farmer targets an exogenously given quantity of wool, but only a random fraction thereof is materialized. The farmers choose between two growing techniques with different yield factor distributions and associated payments. Finally, Mendelson and Tunca (2007) and Pei et al. (2011) are examples of a supply chain model in which the agents can use the spot market as an alternative outlet to buy or sell items; see the references therein.

3. The Model

The manufacturer starts by selecting the set of farmers to contract with, from a set of I potential farmers, who are differentiated by their geographical location, their probability distributions of seasonal rainfall (and surface water, if applicable), their local well capacity as well as production cost and efficiency parameters. Information is *asymmetric* in the sense that some of the parameters are *privately* known to the farmer, and only *distributionally* known to the manufacturer.

Each of the selected farmers *follows* by choosing a contract from a given menu of contracts. The menu may be exogenously specified, or it may be designed endogenously. A contract specifies the guaranteed *unit price* the manufacturer commits herself to pay per unit of supply, as well as the corresponding *quantity* the manufacturer will buy at the end of the season, of course capped by the supply the farmer is able to generate. Typically, a contract with a larger quantity commitment comes with a lower per unit price, thus generating a trade-off for the farmer: if he chooses a contract with a low quantity volume, but higher per unit price, there is a significant chance of oversupply, the excess of which may be sold at the spot market at an uncertain and typically lower price. Conversely, if he selects a contract with a higher quantity volume, the risk of oversupply is reduced, but so is the profit margin.

The menu includes the "no-supply option" which is selected when the maximum expected profit among all offered supply contracts falls below a given *minimum participation threshold* (or when all contracts violate a risk aversion constraint). When evaluating a contract, the farmer takes into account how rainfall and surface water can optimally be supplemented with well water, once uncertainty about the rainfall, surface water, and well capacity have been revealed.

The best response of each farmer to the offered menu of contracts generates a distribution of his supply, were he to be selected, along with the contract choice and associated price-quantity combination. In assessing the supply quantity that would be obtained when selecting any given farmer, the manufacturer faces two sources of uncertainty: first, the farmer, himself, faces a volatile supply quantity due to the intrinsic randomness of various factors (such as rainfall, surface water, and well capacity) determining the yield. The manufacturer's uncertainty is further compounded by the fact that, as mentioned earlier, she has *incomplete*, i.e., only distributional information regarding some of the farmers' parameters. Thus, the manufacturer faces a *mixture* of the possible supply random variables that prevail under given values for the uncertain farmer parameters.

The manufacturer selects a set of farmers to contract with, in order to assure, with a high likelihood, that the aggregate supply is sufficient to cover the aggregate demand which arises at the end of the season at the different manufacturing facilities. (More generally, one may want to trace out the optimal cost - coverage efficient frontier.) This coverage constraint plays a fundamental role in the manufacturer's problem. The manufacturer, typically, needs to secure its supplies from contract farming, for all the various reasons listed in the Introduction. Thus, the spot market is typically only used to cover *limited* shortages –with imperfect products– in the rare scenarios permitted by the probabilistic coverage constraint.

In selecting the set of farmers, the manufacturer aims to minimize along with the expected procurement costs, the expected distribution costs between the farmers and the production facilities. Since the optimal distribution plan can be determined after the supply and demand realizations are observed, the manufacturer faces a two-stage stochastic program, incorporating the farmers' best responses to the offered menu of contracts: in the first stage, the farmers are selected; in the second stage, the distribution plan is determined. The distribution costs are linear in the shipment volumes between the farmers and the facilities. This assumption is satisfied when the shipments either (i) are made in full truckloads, or (ii) are carried out by outside shipping companies. (In our application described in Section 7, both assumptions are satisfied.) Distribution costs often represent a major component of aggregate costs. (In our application, they amounted to at least one third of aggregate costs.) As a consequence, the geographic location of each potential supplier is an important selection criterion, along with many others. This fact is confirmed in our numerical studies; see (41) and (42).

We complete this model section with a formulation of the manufacturer's two-stage stochastic program arising under a *given* menu of contracts. In Section 5, we describe how the design of an optimal menu of contracts can be integrated with this planning problem. We use the subscript i (j) to distinguish among the I farmers (J manufacturing plants). Subscript $i=0$ refers to the spot market (or an alternative emergency source, e.g. an import from a different country), assumed to have an ample supply source S_0.

Parameters and Input Variables:

- D_j = the (random) demand at manufacturing plant j, $j=1,\ldots,J$
- $D^{tot}=\sum_{j=1}^{J} D_j$: the aggregate (random) demand among all manufacturing facilities, with mean μ_{tot} and standard deviation σ_{tot}
- S_i = the (random) supply of farmer i, if selected, resulting from that farmer's best response to the offered menu of contracts, $i=1,\ldots,I$
- v_i = the expected payment to farmer i, if selected, $i=1,\ldots,I$
- γ_{ij} = the distribution cost per unit of crop, dispatched from farmer i to manufacturing facility j, $i=1,\ldots,I$ and $j=1,\ldots,J$
- γ_{0j} = the (random) purchase, transaction and distribution cost per unit, obtained at the spot market (or emergency source) and shipped to facility j, $j=1,\ldots,J$
- ϵ = the maximum permitted probability of an aggregate shortfall at the end of the season

In addition to the expected payments to the selected farmers, captured by the v-vector, there may be fixed transaction costs –e.g., monitoring or quality control costs– incurred for each farmer the manufacturer contracts with. In Appendix K.2, we clarify how such fixed transaction costs can easily be incorporated.

The joint distribution of the demand variables $\{D_j\}$ is an input to the model. The random variables $\{S_i\}$ and parameters $\{v_i\}$ are to be derived from the solution of the farmers' best response problems, as described at the end of Section 4. The distribution cost parameters $\{\gamma_{ij},\ i=1,\ldots,I,\ j=1,\ldots,J\}$ are deterministic cost coefficients, which are inputs to the model. In contrast, the parameters $\{\gamma_{0j},\ j=1,\ldots,J\}$ are random at the first stage of the stochastic program since they involve the end-of-the-season spot price of the commodity. In Appendix H, we discuss several approaches to derive a probability distribution for the latter.

Decision Variables:

- $Y_i=1$ if farmer i is selected, $i=1,\ldots,I$; and 0, otherwise; $Y_0=1$
- ν_{ij} = the volume shipped from source i to manufacturing plant j, $i=0,\ldots,I$ and $j=1,\ldots,J$

The manufacturer's two-stage stochastic program may be formulated as:

$$(M)\qquad \min_{Y\in\{0,1\}^I}\ g(Y)\equiv\sum_{i=1}^{I} v_iY_i+\mathbb{E}_{\{S_i,D_j,\gamma_{0j}\}}\Psi(Y) \tag{1}$$

$$s.t.\ \mathbb{P}\left[\sum_{i=1}^{I} S_iY_i-D^{tot}<0\right]\le\epsilon \tag{2}$$

where the recourse function $\Psi(Y)$ is defined as

$$\Psi(Y) = \min_{\nu} \sum_{i=0}^{I} \sum_{j=1}^{J} \gamma_{ij} \nu_{ij} \tag{3}$$

$$s.t. \sum_{j=1}^{J} \nu_{ij} \leq S_i Y_i \quad i = 0 \ldots I; \quad \sum_{i=0}^{I} \nu_{ij} \geq D_j \quad j = 1 \ldots J; \quad \nu_{ij} \geq 0 \quad i = 0 \ldots I, \; j = 1 \ldots J$$

In the above formulation, it is assumed that any excess in the aggregate supply beyond the aggregate demand, has no salvage value. In Appendix K.1, we show that our analyses can be adapted to allow for such salvage opportunities, via the spot market or other channels. (Similarly, fixed transaction costs per farmer could be added to the model, without any additional complexity; such fixed costs are simply added to the coefficients in the first term of (1).)

In our base model, we assume that the farmers' rainfall (and surface water, if applicable) variables, and hence the $\{S_i\}$ variables are independent of each other, as well as from the aggregate demand variable D^{tot} or the future spot price. Clearly, seasonal rainfall quantities are correlated, in particular among farmers in the same general region as well as with the future spot price. In Appendix K.3, we therefore show how our results can be generalized to allow for such interdependences. Our independence assumption impacts only the specification of the coverage constraint (2). In evaluating the second term in the objective function (1), an arbitrary *joint* distribution of the random variables $\{S_i, D_j, \gamma_{0j}\}$ may be employed.

To appreciate the complexity of the manufacturer's problem (M), it is useful to consider a greatly simplified (yet still NP-hard) version in which all of the manufacturer's decisions may be made under perfect information about all random variables, i.e., the supply volumes $\{S_i\}$ and the demand levels $\{D_j\}$. This *perfect information case* is covered in Appendix I.

The coverage constraint, while essential, adds a significant complication to the farmer selection problem. Note that even a single test whether a given set of farmers, i.e., a given vector Y, satisfies the constraint is very complex and time consuming: it involves the calculation of the convolution of potentially hundreds or even thousands of random variables, each with a complicated distribution, to be determined by an extensive simulation study; see the discussion at the end of Section 4.

Instead, we show that the coverage constraint (2) may be approximated closely by one in which only the first two moments of the random variables $\{S_i\}$ and D^{tot} are needed. More specifically, we replace constraint (2) by:

$$\epsilon \geq \Phi\left(\frac{\mu_{tot} - Y_E}{\sqrt{\sigma_{tot}^2 + \sum_{i=1}^{I} Var(S_i) Y_i}}\right) = 1 - \Phi\left(\frac{Y_E - \mu_{tot}}{\sqrt{\sigma_{tot}^2 + \sum_{i=1}^{I} Var(S_i) Y_i}}\right) \tag{4}$$

where $Y_E = \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i$ and with $\Phi(.)$ the cdf of the standard Normal distribution. This substitution is justified by the following Central Limit Theorem result (proof of this lemma can be found in Appendix A):

LEMMA 1. *For any x,* $\lim_{I\to\infty} \mathbb{P}\left[\sum_{i=1}^{I} S_i Y_i - D^{tot} < x\right] = \Phi\left(\frac{x - Y_E + \mu_{tot}}{\sqrt{\sigma_{tot}^2 + \sum_{i=1}^{I} Var(S_i) Y_i}}\right)$

It is well known that the Normal distribution provides a very good approximation for sums of even a modest number of independent random variables. Note that the number of farmers I is in the hundreds or thousands, so the Normal approximation is virtually exact.

Let z_ϵ be equal to $\Phi^{-1}(1-\epsilon)$. Federgruen and Yang (2008) show that constraint (4) may be replaced by a set of three linear or quadratic constraints. The manufacturer's problem may thus be formulated as:

$$(M') \qquad \min_{Y\in\{0,1\}^I} \; g(Y) \equiv \sum_{i=1}^{I} v_i Y_i + \mathbb{E}_{\{S_i, D_j, \gamma_{0j}\}} \Psi(Y) \tag{5}$$

$$s.t. \; (Y_E - \mu_{tot})^2 - z_\epsilon^2 \sigma_{tot}^2 - z_\epsilon^2 \sum_{i=1}^{I} Var(S_i) Y_i \geq 0 \tag{6}$$

$$Y_E \geq \mu_{tot} + z_\epsilon \sigma_{tot} \tag{7}$$

$$Y_E = \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i \tag{8}$$

4. The Farmer's Best Response Strategy

In any given season, each farmer's yield is random, primarily because of volatility in the water supply. (It is easy to adjust our model and analysis when there are additional or alternative random yield factors.) The latter consists of natural rainfall and surface water, possibly complemented with irrigation water, drawn from a well, if available. Water supply is the primary random yield factor for many crops (e.g., rice, maize, cotton, soybean, and potatoes) and in many countries; see e.g., FAO (2007), Palma (2004), and Dawande et al. (2013).

In many developing countries, the farmers may face a third source of uncertainty beyond the season's rainfall, surface water, and his well capacity: even the availability of an electricity source to enable irrigation from the well may be uncertain; when unavailable, the water may need to be pumped with higher diesel costs. The water volume to be drawn from the well may be determined after the season's actual rainfall and surface water quantity, the well capacity, and the availability of an electricity source become known.

Each farmer is offered a menu of L contracts, indexed by a subscript ℓ, where a contract specifies a unit price and an associated potential purchase quantity by the manufacturer. The actual purchase quantity is the lesser of the potential quantity and the realized yield. If the realized yield at the end of the season exceeds the potential purchase quantity, the farmer can possibly sell the excess quantity on the spot market, where the unit price is (highly) random. In this section, we analyze the farmer's best response problem to a given menu of contracts. In the next section, we show how the best response problem can be used in the design of an *optimal* menu of contracts.

The end-of-the-season spot price is *exogenous* to the farmers' contract selection problem. Consider, as an example, the Indian potato market—the primary input market for our industrial liaison. Total potato production in 2012 – 2013 amounted to well over 40 million tons. As shown in Table 1, the average mid-size farm has less than 3 hectares of cultivated land. Even in the highest yield regions, the average yield is no more than 25 tons per hectare. Thus, even if our manufacturer contracts with 1,000 farmers, the total expected supply is less than 0.2% of the national supply, hence with a negligible impact on the spot price. See *www.potatopro.com/india/potato-statistics.*

	Small	**Medium**	**Large**
Avg. farm size (ha) / Percentage (%)	1.43 / 46.34	2.76 / 31.70	7.95 / 21.95

Table 1 Farm size distribution

We mentioned, in the Introduction, that it is very difficult to predict the spot price several months or a year ahead of its realization. Moreover, farmers, particularly small farmers in developing countries, can ill afford to take significant risks with their livelihood. When considering the spot market as an outlet for any excess supplies, we therefore assume that the farmer considers a safe lower bound for the next year's spot price. (This lower bound may be based on recent price behavior and media based information as to expectations.) However, more sophisticated farming companies may use the actual predicted distribution of the spot price, in lieu of this lower bound. Whatever spot price estimate the farmer is comfortable using, it has an important impact on his contract selection, no less than the salvage value in the classical newsvendor model does.

The farmer's problem is thus specified by the following list of notations.

Parameters and Input Variables for Each Farmer $i = 1, \ldots, I$*:*

- R_i: the random rainfall and surface water, in cubic meters, at farmer i, with cdf $G_i(\cdot)$
- C_i: the random capacity of farmer i's well, in cubic meters
- X_i: farmer i's random yield volume
- S_i : farmer i's random sales volume to the manufacturer
- $\underline{\pi}_i$: farmer i's minimally acceptable profit level, to be met with a minimum likelihood $(1-\epsilon_i)$
- p_ℓ: the per unit purchase price associated with contract ℓ, $\ell = 1, \ldots, L$
- q_ℓ: the potential purchase quantity associated with contract ℓ, $\ell = 1, \ldots, L$
- $\underline{p}_i^s$: the (lower bound) spot price considered by farmer i
- c_i: farmer i's variable production cost, exclusive of irrigation cost
- δ_i^T : farmer i's cost of drawing one cubic meter of water from the well $\in \{\delta_i^L, \delta_i^H\}$, ($\delta_i^L$ is the cost rate when an electricity source is available, δ_i^H is the higher cost rate in the alternative case)

We assume that each of the farmer's parameters $\{\underline{\pi}_i, c_i, \delta_i^L, \delta_i^H\}$ as well as the distributions of the random variables $\{R_i, C_i\}$ are drawn from finitely many types, even as $I \uparrow \infty$.

Decision Variables:

- x_i: the water volume, in cubic meters, drawn from farmer i's well, if available, $i = 1, \ldots, I$
- $Z_{i\ell} = 1$ if farmer i chooses contract ℓ, $i = 1, \ldots, I$ and $\ell = 1, \ldots, L$; and 0 otherwise

X_i, farmer i's yield, is a (non-linear) function of his aggregate water supply, i.e.,

$$X_i = f_i(\alpha_i, T_i) \qquad \text{with } T_i = R_i + x_i \tag{9}$$

where α_i denotes a parameter that is privately known to the farmer. The manufacturer has a prior distribution for the parameter value, with cdf $F_i(\cdot)$. For notational simplicity, we assume that the information asymmetry pertains to a single parameter α_i. Generalizations to settings where multiple parameters in the yield function $f_i(\cdot,\cdot)$ are privately known, can be readily accommodated.

We make the following assumption regarding the shape of the yield function.

Assumption 1. *The yield function $f_i(\cdot, T_i)$ is concave in $T_i = R_i + x_i$.*

The following are four examples of yield functions that satisfy Assumption 1.

Structure 1: α_i represents the ideal yield under optimal water supply conditions. The yield X_i is a linear function of the shortfall of the water supply vis-a-vis an ideal water quantity W_i, i.e.,

$$X_i = \alpha_i - \beta(W_i - R_i - x_i)^+ \tag{10}$$

Structure 2: Identical to Structure 1, except that the yield declines (linearly) if the water supply is in excess of the ideal water quantity W_i, i.e.,

$$X_i = \alpha_i - \beta^-(W_i - R_i - x_i)^+ - \beta^+(R_i + x_i - W_i)^+ \tag{11}$$

Structure 3: Identical to Structure 1, except that the maximum yield α_i is maintained as long as the aggregate water supply is between W_i and $\widehat{W}_i$, and declines linearly thereafter, i.e.,

$$X_i = \min\{\alpha_i, \alpha_i - \beta^-(W_i - R_i - x_i), \alpha_i - \beta^+(R_i + x_i - \widehat{W}_i)\} \tag{12}$$

Structure 4: This structure has been suggested in the agricultural literature, see e.g., Christensen and McElyea (1988), Rao et al. (1990), Palma (2004), and Dawande et al. (2013).

$$X_i = \alpha_i(1 - \exp(-\beta_i(R_i + x_i))) \tag{13}$$

Equation (9) represents X_i, the farmer i's yield, as a function of the *aggregate* water supply. In reality, X_i may depend on the aggregate of the *daily* shortfalls and surpluses vis-a-vis an ideal daily water quantity, and therefore depend on the stochastic process of *daily* rainfalls. More specifically, let R_{it} be the rainfall during day t at farmer i and assume the growing season consists of T days, numbered $t = 1,\ldots,T$. Then, under Structure 1, for example, the yield equation (10) should be replaced by

$$X_i = \alpha_i - \beta \sum_{t=1}^{T} \left[\frac{W_i}{T} - R_{it} - \frac{x_i}{T} \right]^+ \leq \alpha_i - \beta \left(W_i - R_i - x_i \right)^+ \tag{14}$$

Many stochastic processes have been proposed to represent the rainfall process $\{R_{it}\}$, see e.g., Waymire and Gupta (1981) and Rodriguez-Iturbe et al. (1987). While we will proceed with the aggregate yield model (9), its replacement by (14) (or other such yield models) is easily accommodated. Note that, $S_i = \min\left\{ \sum_{\ell=1}^{L} q_\ell Z_{i\ell},\ X_i \right\}$, $i = 1,\ldots,I$.

Let $\Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T)$ denote the random profit earned by farmer i when choosing contract ℓ:

$$\Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T) \equiv \max_{x_i} \left\{ (p_\ell - c_i)\min\{q_\ell, f_i(\alpha_i, R_i + x_i)\} + (\underline{p}_i^s - c_i)(f_i(\alpha_i, R_i + x_i) - q_\ell)^+ - \delta_i^T x_i \right\} \tag{15}$$

$$s.t.\ 0 \leq x_i \leq C_i$$

Each farmer i's best response problem may be formulated as the following two-stage stochastic program.

$$(F) \quad \max_{Z_{i\ell}} \sum_{\ell=1}^{L} Z_{i\ell} \left[\mathbb{E}_{\{R_i, C_i, \delta_i^T\}} \Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T) \right] \tag{16}$$

$$s.t. \sum_{\ell=1}^{L} Z_{i\ell} \leq 1 \tag{17}$$

$$\sum_{\ell=1}^{L} Z_{i\ell}\ \mathbb{P}_{\{R_i, C_i, \delta_i^T\}} \left[\Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T) < \underline{\pi}_i \right] \leq \epsilon_i \tag{18}$$

In other words, we formulate the farmer's problem as selecting the contract which maximizes his expected profits subject to a risk aversion constraint. For the latter, we choose a maximum probability ϵ_i with which the farmer's profits fall below the minimum profit level $\underline{\pi}_i$. Constraint (18) may be replaced or complemented with an upper bound on the standard deviation of the profit level $\left(\sum_{\ell=1}^{L} Z_{il} \left[\text{stdev}_{\{R_i, C_i, \delta_i^T\}} \Pi_{i\ell}(R_i, C_i, \delta_i^T) \right] \right)$ or any other risk measure.

We now show how, for any of the contracts $\ell = 1,\ldots,L$, the profit value $\Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T)$ may be efficiently evaluated for any given quartet $(\alpha_i, R_i, C_i, \delta_i^T)$. Recall that the farmer has full knowledge of the parameters in (15), including the structural form of the function $f_i(\alpha_i, \cdot)$.

Under Assumption 1, $f_i(\alpha_i, T_i)$ is concave, hence differentiable in T_i everywhere, with the possible exception of a countable set of values, where, in any case, the left and right hand partial derivatives $\partial^- f_i/\partial T_i$ and $\partial^+ f_i/\partial T_i$ exist. Let $\phi_{i\ell}(\alpha_i, T_i) \equiv (p_\ell - c_i)\min\{q_\ell, f_i(\alpha_i, T_i)\} + (\underline{p}_i^s - c_i)(f_i(\alpha_i, T_i) - q_\ell)^+$ and note that the optimization problem (15) can be written as

$$\Pi_{i\ell}(\alpha_i, R_i, C_i, \delta_i^T) = \max_{0 \leq x_i \leq C_i} \left\{ \phi_{i\ell}(\alpha_i, R_i + x_i) - \delta_i^T x_i \right\} \tag{19a}$$

$$= \max_{R_i \leq T_i \leq R_i + C_i} \left\{ \phi_{i\ell}(\alpha_i, T_i) - \delta_i^T T_i \right\} + \delta_i^T R_i \tag{19b}$$

LEMMA 2. *Assume the yield function $f_i(\cdot,\cdot)$ satisfies Assumption 1.*

(a) Let $T^*_{i\ell}(\alpha_i, R_i)$ denote the smallest optimizer of (19b). Then

$$T^*_{i\ell}(\alpha_i, R_i) = \min\left\{R_i + C_i,\ \min\left[T_i \geq R_i :\ \partial^-\phi_{i\ell}(\alpha_i, T_i)/\partial T_i \leq \delta^T_i\right]\right\}.$$

(b) $T^*_{i\ell}(\alpha_i, R_i)$ is increasing in R_i.

(c) If the yield function $f_i(\cdot,\cdot)$ is supermodular, as in all structures 1-4, then $T^*_{i\ell}(\alpha_i, R_i)$ is increasing in α_i.

(d) Assume the yield function $f_i(\alpha_i, T_i)$ is strictly increasing and differentiable in T_i, then $T^*_{i\ell}(\alpha_i, R_i)$ may be determined as follows:

Let $t^*_i(\alpha_i, q_\ell)$ denote the water supply quantity which generates a yield equal to the contract volume q_ℓ, i.e., $t^*_i(\alpha_i, q_\ell)$ is the unique root of the equation

$$f_i(\alpha_i, T_i) = q_\ell,\ \textit{if}\ \lim_{T_i\to\infty} f_i(\alpha_i, T_i) \geq q_\ell,\ \textit{and}\ t^*_i = \infty,\ \textit{otherwise}. \tag{20}$$

$$\textit{If}\ t^*_i < \infty, \quad \partial t^*_i(\alpha_i, q_\ell)/\partial q_\ell = [\partial f_i(\alpha_i, t^*_i)/\partial T_i]^{-1}. \tag{21}$$

(I) $R_i \leq t^*_i(\alpha_i, q_\ell)$:

(i) If $(p_\ell - c_i)\partial f_i(\alpha_i, R_i)/\partial T_i \leq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = R_i$.

(ii) If $R_i + C_i \leq t^*_i(\alpha_i, q_\ell)$ and $(p_\ell - c_i)\partial f_i(\alpha_i, R_i + C_i)/\partial T_i \geq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = R_i + C_i$.

(iii) If $R_i + C_i > t^*_i(\alpha_i, q_\ell)$ and $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, R_i + C_i)/\partial T_i \geq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = R_i + C_i$.

(iv) If (i)-(iii) do not apply and $(p_\ell - c_i)\partial f_i(\alpha_i, t^*_i)/\partial T_i < \delta^T_i$, then $T^*_{i\ell}$ is the unique root of the equation $(p_\ell - c_i)\partial f_i(\alpha_i, T_i)/\partial T_i = \delta^T_i$ on the interval $[R_i, t^*_i]$.

(v) If (i)-(iii) do not apply and $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, t^*_i)/\partial T_i > \delta^T_i$, then $T^*_{i\ell}$ is the unique root of the equation $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, T_i)/\partial T_i = \delta^T_i$ on the interval $[t^*_i, R_i + C_i]$.

(vi) If (i)-(iii) do not apply and $(p_\ell - c_i)\partial f_i(\alpha_i, t^*_i)/\partial T_i \geq \delta^T_i$ and $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, t^*_i)/\partial T_i \leq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = t^*_i$.

(II) $R_i > t^*_i(\alpha_i, q_\ell)$:

(i) If $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, R_i)/\partial T_i \leq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = R_i$.

(ii) If $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, R_i)/\partial T_i \geq \delta^T_i$, then $T^*_{i\ell}(\alpha_i, R_i) = R_i + C_i$.

(iii) Otherwise, $T^*_{i\ell}$ is the unique root of the equation $(\underline{p}^s_i - c_i)\partial f_i(\alpha_i, T_i)/\partial T_i = \delta^T_i$ on the interval $[t^*_i, R_i + C_i]$.

Lemma 2(d) provides an analytical characterization of $T^*_{i\ell}(\alpha_i, R_i)$, the farmer's optimal total water supply when adopting contract ℓ from the menu, and given a rainfall value R_i. This optimal water supply value is either given as a *constant* or as the unique root of a simple monotone function. The most general case where $f_i(\alpha_i, \cdot)$ may fail to be increasing or differentiable everywhere, can be handled analogously, except that more cases need to be distinguished. Note also that the dependence of $T^*_{i\ell}$ on q_ℓ is very simple: $\partial T^*_{i\ell}/\partial q_\ell = 0$ in all but case $(I)(vi)$ where it is given by the simple expression in (20).

The expected profit under each of the L contracts, as well as the likelihood of this profit value meeting the minimum level $\underline{\pi}_i$, can thus be evaluated by unconditioning over the joint distribution of the natural rainfall (and surface water, if applicable), the well capacity, and a binary variable indicating whether a local electricity source is available. The farmer then chooses, among all contracts which satisfy the specified risk bound, the one with the highest expected profit value. (If none of the contracts are feasible, farmer i rejects the complete menu of contracts.) Under the optimal contract, the distribution of the supply quantity $S_i(\alpha_i)$ can be evaluated, most simply, by generating a sample of K realizations from the above joint distribution of the random variables.

We conclude this section with a specification of the random variables $\{S_i\}$ and the expected payment values $\{v_i\}$ which are required for the specification of the manufacturer's problem (1)-(3): The $\{S_i\}$ variables can be constructed as a mixture of the $\{S_i(\alpha_i)\}$ variables above, with the cdf F_i as the mixing distribution. Finally,

$$v_i = \int p^{\alpha}_{(i)} \mathbb{E}(S^{\alpha}_i) dF_i(\alpha),$$

with $p^{\alpha}_{(i)}$ the unit price in farmer i's optimal contract, assuming farmer i faces a specific parameter (vector) α_i.

It is clear that, under asymmetric information, the supply variables are more volatile than when information is symmetric. It stands to reason that this results in a larger pool of contracted farmers and a larger expected aggregate supply value, since informational uncertainty now compounds on the intrinsic supply and demand risks. Our numerical studies confirm this conjecture.

5. Selecting the Menu of Contracts

In the previous section, we analyzed the farmer's best response problem to an exogenously given menu of contracts. In this section, we show how this menu of contracts may be selected endogenously by the manufacturer. In doing so, we follow a standard paradigm in contract theory, see e.g., Bolton and Dewatripont (2005) or Lovejoy (2006), however applied to a complex operations setting.

Typically, a single uniform menu needs to be offered to all farmers in a broad category, for example all farmers in a given district that are of medium size, say, see Table 1.

For notational simplicity only, we formulate the menu design problem assuming a single menu is offered to all farmers. The generalization to multiple menus –one offered to each of several broad categories of farmers– is straightforward.

Assume, therefore, that there are K *types* of farmers, differentiated by their privately known parameter value α_i. In other words, there is a list of K α-values, $\alpha^1 < \alpha^2 < \ldots < \alpha^K$, such that, for all i, $\alpha_i = \alpha^k$ for some k. The manufacturer's prior knowledge about each farmer i's parameter value α_i is captured by a discrete cdf $F_i(\cdot)$ with support on $\{\alpha^1, \ldots, \alpha^K\}$. A menu of K contracts $\{(p_\ell, q_\ell): \ \ell = 1, \ldots, K\}$ is to be designed such that, in accordance with the revelation principle, farmers of type k are incentivized to select the k^{th} contract. The K contracts are to be chosen on a decreasing curve $p = P(q)$. Without loss of generality, the contracts are numbered such that $q_1 \leq q_2 \leq \ldots \leq q_K$ and hence $p_1 = P(q_1) \geq p_2 = P(q_2) \geq \ldots \geq p_K = P(q_K)$.

When specifying the menu design problem, we address a somewhat simplified version of the manufacturer's problem, where in the objective (1) of (M) (or (5) of (M')) the second, distribution cost related term is omitted, as are any risk bounds in the farmer's problem, see (18). With the menu specified as the optimal solution to this contract design program, below referred to as (M^{cont}), the manufacturer solves the complete integrated farmer selection and distribution planning problem based on the anticipated fully optimal contract choices of the farmers that incorporate their risk bounds. In other words, the manufacturer solves the complete problem (M') without any simplifications (or relaxations). In doing so, the manufacturer uses the methods described in Section 6 and Appendix B.

For the sake of notational simplicity, we treat the well capacities $\{C_i\}$ and the water drawing cost rates $\{\delta_i^T\}$ as deterministic. Let

- $U_i(\ell|k,q)$: the expected profit value for farmer i assuming he is of type k and selects contract ℓ; $1 \leq k, \ell \leq K$.
- $\mathbb{E}(S_i|q)$: the expected supply volume for farmer i, $i = 1, \ldots, I$, if selected.
- $Var(S_i|q)$: the variance of the supply volume for farmer i, $i = 1, \ldots, I$, if selected.
- $v_i(q)$: the expected payment to farmer i, $i = 1, \ldots, I$, if selected.
- $\underline{\pi}_{ik}$: the minimally acceptable profit level for farmer i assuming he is of type k; $i = 1, \ldots, I$, $1 \leq k \leq K$.

Treating the volume quantities $\{q_1, \ldots, q_K\}$ as decision variables, we respecify the manufacturer's problem (M') as:

$$(M^{\text{cont}}) \qquad \min_{Y \in \{0,1\}^I, q} \sum_{i=1}^{I} v_i(q) Y_i \tag{22}$$

$$s.t. \ (Y_E - \mu_{tot})^2 - z_\epsilon^2 \sigma_{tot}^2 - z_\epsilon^2 \sum_{i=1}^{I} Var(S_i|q) Y_i \geq 0 \tag{23}$$

$$Y_E \geq \mu_{tot} + z_\epsilon \sigma_{tot} \tag{24}$$

$$Y_E = \sum_{i=1}^{I} \mathbb{E}(S_i|q) Y_i \tag{25}$$

$$q_1 \leq q_2 \leq \ldots \leq q_K \tag{26}$$

$$U_i(k|k, q_k) Y_i \geq \underline{\pi}_{ik} Y_i, \qquad i = 1, \ldots, I \text{ and } k = 1, \ldots, K \qquad \text{(IR)} \tag{27}$$

$$U_i(\ell|k, q_\ell)Y_i \leq U_i(k|k, q_k)Y_i, \qquad i = 1, \ldots, I, \quad k \neq \ell \qquad \text{(IC)} \tag{28}$$

Constraint set (27) represents each farmer's *individual rationality* constraints (IR) ensuring that his expected profit meets the minimum participation level. Constraint set (28) represents the *incentive compatibility* constraints (IC) ensuring that farmer i selects contract k if of type k. Appendix C provides analytical expressions for the various quantities $U_i(\ell|k, q)$, $\mathbb{E}(S_i|q)$, $Var(S_i|q)$, and $v_i(q)$ required in the model formulation.

In some contract design problems, it is possible to replace the *full* set of (IC) –incentive compatibility constraints (28)– by a subset of so-called "*local IC constraints*", where for any $\ell = 1, \ldots, K$ only the constraints corresponding with the values $\ell = k + 1$ and $\ell = k - 1$ need to be incorporated, see e.g., Bolton and Dewatripont (2005) or Lovejoy (2006). This reduction of the set of IC-constraints would be possible if the maximand in the farmer's objective function (19a) were supermodular in (q_ℓ, α_i). However, the maximand fails to be supermodular, even if the yield function $f_i(\alpha_i, T_i)$ is.

The menu design problem (M^{cont}) is a highly non-linear mixed integer problem. The natural solution method is to alternate between

- (M_q) identifying the optimal set of farmers, i.e., the optimal Y-vector, for a given menu of contracts, i.e., a given q-vector, and
- (M_Y) identifying an optimal contract menu, i.e., an optimal q-vector, for a given set of selected farmers, i.e., a given Y-vector.

This iterative algorithm is guaranteed to converge in finitely many iterations; if completed till convergence, it generates an optimal solution.

Problem (M_q) is an integer problem, and a special case of the problem discussed in Section 6 and Appendix B; any of the methods proposed there can be applied to solve this problem. Problem (M_Y), on the other hand, is a continuous non-linear program with a small number (K) of continuous variables, but $O(IK^2)$ constraints. With the exception of the supply coverage constraints (23) – (25), all other constraints, as well as the objective function are separable functions. All of the (IR) and (IC) constraints, (27) and (28), involve only one or two of the decision variables.

6. The Manufacturer's Problem: Approximate Solution Methods

In this section, we develop approximate solution methods for the manufacturer's problem (M') under a *given* menu of contracts. These fall into two broad categories covered in Subsections 6.1 and 6.2, respectively.

a) *Constructive improvement heuristics*: These are iterative methods in which, in each iteration, the current set of suppliers is modified by the addition or elimination of a supplier or the replacement of a supplier by one outside the current set. The method terminates if none of the relevant set perturbations results in a cost improvement.

b) *Mathematical programming based heuristics*

In Appendix B, we derive a lower bound for (M') via the Lagrangean dual of one of its formulations; moreover we show how this Lagrangean dual can be computed efficiently.

6.1. Constructive Improvement Heuristics

The formulation of the stochastic program (M) in (1)-(3), or its approximation (M') in (5)-(8), projects the optimization problem onto $\{0,1\}^I$, the space of selection variables $\{Y_i\}$. The vector Y specifies the set of selected farmers. We first show that the objective function $g(Y)$ in (1) is a *supermodular* set function. This means that for any pair of vectors $Y^1 < Y^2$,

$$g(Y^2 + e_i) - g(Y^2) \geq g(Y^1 + e_i) - g(Y^1)$$

with e_i the i^{th} unit vector in R^I. In other words, the addition of a new farmer to a given set of farmers results in a smaller expected cost saving, hence, a larger expected cost increase, compared to when the same farmer is added to a subset thereof. Appendix A shows the proof of Theorem 1.

Theorem 1. *The objective function of problem (M) is a supermodular set function.*

Supermodular set functions are important in the optimization and economics literature. For example, many fundamental problems in a variety of supply chain models, including network design problems, stochastic inventory problems, and pricing models can be posed as a problem of minimizing a supermodular set function. (An example is the uncapacitated plant location problem, see Cornuejols et al. 1977). This class of problems -even unconstrained- was shown to be NP-hard even if the evaluation of the set function can be performed in polynomial time, as is the case in our model, where it reduces to the solution of a series of independent transportation problems. The complexity results follow from the fact that the general class of problems generalizes many well-known NP-hard problems, including the special case discussed in Appendix I (with $K = 1$ scenarios) and the aforementioned uncapacitated plant location problem.

However, Cornuejols et al. (1977) and Nemhauser and Wolsey (1978) showed that, at least for the *unconstrained* (supermodular) set minimization problem, simple heuristics such as the greedy-add or greedy-drop heuristics work remarkably well. (The simplest version of the greedy-add heuristic starts from a feasible solution and adds, at each iteration, a single new element that results in the largest cost saving among all possible choices; the heuristic terminates when no addition of any outside element results in a cost decrease. The greedy-drop heuristic starts with the full set of elements, and in each iteration, drops whatever element results in the smallest cost increase.) In numerical studies, these heuristics have performed remarkably well. Moreover, Cornuejols et al. (1977) derives an impressive worst case optimality gap of $e^{-1} \approx 37\%$, albeit that a somewhat unconventional optimality gap measure is used. More recently, the theory has been generalized to include side constraints that are linear in the binary variables that describe the composition of the selected set; see Lee et al. (2009) and Kulik et al. (2013).

Because of the fixed costs of engaging with the farmers, i.e., the first term in (1), $g(Y)$ is not a monotone function of the selected set of farmers, which makes solving this problem more complex (see Kulik et al. 2013). Nevertheless, these results bode well for the performance of similar constructive heuristics for our problem, in particular:

Greedy-Add Heuristic, version 1: In this heuristic, we add, in each iteration, a new supplier to the current set of suppliers. As long as the coverage constraint fails to be satisfied, we add any supplier whose addition results in the smallest expected cost increase; in subsequent iterations, we execute the addition only if the resulting expected cost increment is *negative*, and terminate otherwise. As the initial set of suppliers, we select the solution to the perfect information problem (PIM), see Appendix I, with a single scenario $(K = 1)$ and all random variables $\{S_i\}$ and $\{D_j\}$ replaced by their expected values.

The determination of the initial set of suppliers reduces to the solution of a fairly simple Mixed Integer Linear Program $(MILP)$. The complexity of the subsequent iterations amounts to at most $O(I^2)$ evaluations of the set function $g(\cdot)$ and the same number of tests whether the coverage constraint is satisfied. Appendix J presents two variants of this greedy heuristic. Their performances are comparable to the *Greedy-Add heuristic, version 1*, but, in our numerical experiments, are dominated by the latter.

6.2. A Mixed Integer Programming Method

The manufacturer's problem (M'), i.e., the problem of minimizing (5), subject to (6) - (8) would reduce to a Mixed Integer Linear Program $(MILP)$, and thereby solvable for instances with thousands of binary variables, were it not for the single quadratic constraint (6). This suggests a heuristic approach where the quadratic term $(Y_E - \mu_{tot})^2$ in (6) is replaced by a piecewise linear approximation. Note that by (7) and (8):

$$d_0 \equiv \mu_{tot} + z_\epsilon \sigma_{tot} \leq Y_E = \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i \leq \sum_{i=1}^{I} \mathbb{E}(S_i) \equiv d_N$$

Our $MILP$-based heuristic thus consists of selecting a grid $\{d_0, d_1, \ldots, d_N\}$ on the feasible interval $[d_0, d_N]$, replacing the quadratic function $h(Y_E) = (Y_E - \mu_{tot})^2$ by a piecewise linear function $\tilde{h}(Y_E)$ which coincides with the quadratic function $h(.)$ at the grid points and is linear in between. Thus, representing Y_E as a weighted average of the grid points, with the help of the auxiliary variables $\{\eta_n : n = 0, 1, \ldots, N\}$, we replace problem (M') by the $MILP$:

$$\min_{Y \in \{0,1\}^I} \; g(Y) \equiv \sum_{i=1}^{I} v_i Y_i + \mathbb{E}_{\{S_i, D_j, \gamma_{0j}\}} \Psi(Y) \tag{29}$$

$$
\begin{aligned}
s.t. \; & \sum_{n=0}^{N} \eta_n (d_n - \mu_{tot})^2 - z_\epsilon^2 \sigma_{tot}^2 - z_\epsilon^2 \sum_{i=1}^{I} Var(S_i) Y_i \geq 0 \\
& \sum_{n=0}^{N} \eta_n = 1; \quad \sum_{n=0}^{N} \eta_n d_n = \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i
\end{aligned}
$$

along with so-called SOS2 constraints to ensure that at most two of the η-variables are nonzero and that these non-zero variables have consecutive indices. As before, we employ the sample averaging method to evaluate the second term in (29).

The full formulation of the $MILP$ is thus given by:

$$
\begin{aligned}
\min_{Y,\nu,\eta} \; & \sum_{i=1}^{I} v_i Y_i + \sum_{k=1}^{K} \sum_{i=0}^{I} \sum_{j=1}^{J} \delta_k \gamma_{ij} \nu_{ijk} \qquad (30) \\
s.t. \; & \sum_{n=0}^{N} \eta_n (d_n - \mu_{tot})^2 - z_\epsilon^2 \sigma_{tot}^2 - z_\epsilon^2 \sum_{i=1}^{I} Var(S_i) Y_i \geq 0 \\
& \sum_{n=0}^{N} \eta_n = 1; \quad \sum_{n=0}^{N} \eta_n d_n = \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i \\
& \sum_{j=1}^{J} \nu_{ijk} \leq S_{ik} Y_i \quad i = 0 \dots I \quad k = 1 \dots K; \quad \sum_{i=0}^{I} \nu_{ijk} \geq D_{jk} \quad j = 1 \dots J \quad k = 1 \dots K \\
& \nu_{ijk} \geq 0, \; Y_i \in \{0,1\}, \; \eta \in SOS2
\end{aligned}
$$

where δ_k, S_{ik}, D_{jk}, and ν_{ijk} are defined as in Appendix I. This program is linear in all the decision variables Y, ν, and η. CPLEX, as well as many other optimization software systems, have built-in functions to represent SOS2 constraints.

7. Numerical Analysis

We have applied our model to a large global manufacturer of potato chips contracting with thousands of potato farmers in India. More specifically, we have created a numerical study with instances that are anchored on data provided by the manufacturer, as well as data that are publicly available. In Appendix D, we describe in detail how the problem instances were created. The exact numbers are withheld in order to preserve confidentiality.

We have evaluated problem instances with seven different values for the *targeted* total number of potential farmers I: I =50, 100, 150, 250, 500, and 1,000. These farmers are distributed over the set of 19 districts, in six of India's states, in which the manufacturer contracted during the two years (2009 and 2010) used for our analysis. Appendix D describes the precise procedure employed to distribute the total number of farmers, I, over the 19 districts listed in Table 10 in Appendix G. For each value of I, we have evaluated 20 problem instances by considering five problem instances for each of four demand level / coverage level combinations: *high* and *low* demand scenarios at a coverage level of 95% and 99%, i.e., $\epsilon = 0.05$ and $\epsilon = 0.01$, respectively. These five problem instances use different realizations of the random parameters that specify an instance: the precise distribution of the I farmers over the 19 districts, the farm sizes assigned to the farmers, as well as the sample of realizations drawn from the joint supply and demand volume distribution $\{S_i, D_j\}$; see (1)-(3).

In Table 2, we report the performance of the Greedy-Add heuristic (version 1 –the other greedy heuristics presented in Appendix J were generally dominated by this *Greedy-Add* procedure) and the mixed integer programming approach, described in Sections 6.1 and 6.2, respectively. (We used 5e-4 as the relative $MILP$ gap / optimality tolerance for the $MILP$ method and when solving the perfect information problem to determine the initial set of farmers in the Greedy-Add heuristic.) Table 2 reports the average of the performance measures across the five instances that were generated for each of the demand level / coverage level combinations. By computing the Lagrangean dual as described in Appendix B, we determined an upper bound for the optimality gap of the heuristic methods.

				Greedy-Add				**MILP**			
TPFS	DL	CL	R_1	RT	CNF	OG	R_2	RT	CNF	OG	R_2
50	H	99%	1.386	2.6	42.4	1.23%	1.238	111.9	43.4	0.63%	1.232
		95%	1.400	1.6	40	1.62%	1.175	5.9	39.6	0.54%	1.165
	L	99%	1.873	2.7	36.8	2.48%	1.245	61.0	36.8	1.09%	1.231
		95%	1.832	1.8	28.4	2.37%	1.172	17.3	29.2	0.70%	1.165
100	H	99%	1.374	4.9	88.8	0.81%	1.229	496.7	88.4	0.27%	1.225
		95%	1.445	10.3	81	1.13%	1.160	804.2	83.6	0.26%	1.158
	L	99%	1.849	10.9	70.8	1.33%	1.229	251.6	69.8	0.63%	1.224
		95%	1.977	8.9	54.8	1.96%	1.162	66.2	52.8	0.36%	1.156
150	H	99%	1.548	26.5	124.2	0.57%	1.225	47.8	124.8	0.28%	1.223
		95%	1.441	15.7	121.8	0.65%	1.158	94.7	118.6	0.19%	1.156
	L	99%*	1.783	28.0	104.2	1.00%	1.225	130.6	102.2	0.40%	1.222
		95%	1.907	22.6	85.4	1.00%	1.158	306.7	86.6	0.34%	1.156
250	H	99%	1.476	89.0	217	0.40%	1.223	50.3	213.8	0.14%	1.221
		95%	1.419	55.4	199.4	0.72%	1.156	136.3	204.2	0.20%	1.155
	L	99%	1.834	134.2	169.4	0.81%	1.221	199.2	161.4	0.29%	1.219
		95%	1.766	66.5	161.2	1.11%	1.156	136.5	162	0.19%	1.154
500	H	99%	1.437	482.7	445	0.24%	1.221	263.1	424.6	0.08%	1.219
		95%	1.430	354.4	392.8	0.59%	1.155	266.5	396.6	0.08%	1.154
	L	99%	1.904	847.1	321	0.59%	1.218	269.3	317.4	0.17%	1.217
		95%**	1.879	469.5	270.6	1.21%	1.155	484.3	290	0.10%	1.153
1,000	H	99%	1.455	23,723.5	866	0.28%	1.219	595.5	814.2	0.18%	1.219
		95%	1.438	15,217.8	758.2	0.45%	1.154	412.7	775.2	0.08%	1.154
	L	99%	1.926	34,253.6	617.8	0.59%	1.217	408.6	616.4	0.21%	1.216
		95%	1.875	14,941.7	528.4	0.86%	1.153	406.7	580.2	0.08%	1.152

Table 2 **Performance of** *Greedy-Add* **and** $MILP$ **heuristics**

TPFS: Targeted Potential Farmer Size, DL: Demand Level, CL: coverage level, RT: Runtime (sec), CNF: Contracted Number of Farmers, OG: Optimality Gap

Column 4 reports the supply-demand ratio R_1, defined as $\sum_{i=1}^{I} \mathbb{E}(S_i)$, the expected supply if all I farmers were contracted, divided by the expected aggregate country-wide demand of the manufacturer. Columns 5-8 and 9-12 report four performance measures for the above two heuristic methods: (a) the runtime in CPU seconds, (b) the number of farmers contracted, (c) the upper bound for the optimality gap, defined as

$$\text{Optimality Gap} = \frac{\text{the expected cost of the heuristic solution} - \text{the value of the Lagrangean dual}}{\text{the value of the Lagrangean dual}},$$

and (d) a supply / demand ratio R_2, defined as $(\sum_{i=1}^{I} \mathbb{E}(S_i)Y_i)/\sum_{j=1}^{J} \mathbb{E}(D_j)$. Thus R_2 reflects the expected *contracted* supply, as opposed to the expected *total* supply from all potential farmers, as in R_1.

We conclude that both solution methods are remarkably close to optimal and that the Lagrangean dual generates a remarkably close-to-accurate estimate of the optimal cost value. The average optimality gap for the Greedy-Add heuristic is 1.09% and that for the $MILP$ based method is 0.35%. Indeed, the latter performs uniformly better. Moreover, the worst case optimality gap across all scenarios is 5.02% for the Greedy-Add procedure and 1.53% for the $MILP$ based approach. Even more encouraging is the fact that the optimality gap decreases with I, to zero, suggesting that both heuristics are asymptotically optimal. Clearly, the results also imply that the average accuracy gap of the Lagrangean dual based lower bound is no more than 0.35%, as well.

The Greedy-Add heuristic is faster, when $I < 250$. When $I = 250$ or $I = 500$, the CPU times are comparable. However, for $I = 1,000$, the $MILP$ method is significantly faster. Most importantly, even instances with $I \sim 1,000$ can be solved in a few CPU minutes.

The extremely low optimality gap for the Greedy-Add heuristic (1.09%, on average) outdoes the expectations raised by Theorem 1 and the subsequent discussion in Section 6. Beyond confirming that this simple heuristic is a competitive alternative solution method, it also indicates that robust farmers' selections can be made in an expanding environment: if the buyer's demand volumes grow from one season to the next, so that the optimal number of contracted farmers increases from I^1 to I^2, very little is lost by adding $(I^2 - I^1)$ new farmers to the initial (close-to-) optimal set of I^1 farmers. Thus, absent major changes in the model parameters, the contracted farmer base may be gradually expanded, without any significant optimality loss, allowing for stability and loyalty towards the selected farmer pool.

The results show that with a modest number of suppliers ($I = 50$), the supply-to-demand ratio R_2 is approximately 1.23 when pursuing a coverage level of 99%. The safety margin can be expected to be significantly larger when, in contrast to our scenarios, there is more significant demand volatility, or as the degree of information asymmetry is larger. Clearly, the required safety margin is larger for scenarios with a coverage level of 99% as opposed to those with a coverage level of 95%. For the latter, the required safety margin varies between 17% and 15% as the number of potential farmers I is increased.

An interesting question is whether one or a few farmer characteristics can be identified that determine whether a farmer is selected by the manufacturer or not. To answer this question, we explored the instances associated with the categories marked with a $*$ and $**$ in Table 2, in detail. Of the many considered farmer characteristics, the following three stood out when running logistic regressions: (i) the expected payment to the farmer divided by his expected supply volume ($v_i/\mathbb{E}(S_i)$), (ii) the coefficient of variation of his supply volume ($CV(S_i)$), and (iii) his distance to the nearest manufacturing plant (Δ_i). See Appendix E for a specification of the logit model and its estimated coefficients.

While insightful, the findings also show that the selection problem involves complex trade-offs among several farmer characteristics. Our Stackelberg game and the associated mathematical program (M') efficiently determine the best among all 2^I subsets of the I farmers. In contrast, assume we confined ourselves to the I sets with the k lowest ($v_i/\mathbb{E}(S_i)$) values, $k=1,\ldots,I$, and identified the best such feasible set. 16% of the (close-to-) optimal farmer set found by (M') fails to be part of this heuristic set.

The above instances assume, as in our base model, that the supply random variables $\{S_i\}$ are independent and that the $\{\gamma_{0j}\}$-coefficients are independent of the aggregate rainfall. In Appendix K.3, we have explained how our model and solution methods can be extended when either one of these two independence assumptions is relaxed. We have investigated how correlations among the $\{S_i\}$ variables impact the optimal choices for the manufacturer. (For a given menu of contracts, these correlation structures do not affect the farmers.) More specifically, we have re-evaluated the eight sets of problem instances with $I=250$ and $I=500$ farmers, now assuming that the rainfall quantities and the $\{\alpha_i\}$ (intercept in the yield equation) parameters for farmers in a given state are drawn from a multivariate Normal distribution with the same means and variances as in the base model, but a common correlation coefficient $\kappa = 0.25$, 0.5, and 0.75. The main managerial conclusions from this study are summarized in Section 8, under Conclusion (j).

Impact of the Menu Design

We have generated a small second set of scenarios to illustrate how our model can be used to evaluate alternative contract menus and how the magnitude of various risk factors impacts the system performance. To this end, we used a base scenario, with a coverage level of 95% and a "high" demand scenario, targeting $I=150$ potential farmers, this time assuming all farmers are small-size farmers; see Table 1. (Because of the above described distribution procedure across the 19 districts, the actual number of potential suppliers $I=158$.)

Here, we report on six sets of exogenously specified contract menus. Appendix F reports on a parallel study in which contract menus are determined *endogenously*, based on the methodology of Section 5.

The six sets of menus are differentiated by (i) the number of contract options: three vs. five; (ii) the slope of the menu line: -10/3, -5/3, and -2/3. Each set consists of six menus, one for each state. As in our first numerical study (see Table 2), all contracts pertaining to the same menu are characterized by price-quantity pairs that lie on a line. See, however, below for a study where the contracts are positioned on a non-linear curve. In each state, all six menus share the same central contract (p^s, $1.43q^s$), as the central contract choice in the large scale numerical study reported in Table 2; see Appendix D and Table 11 in Appendix G for details. In each of the menus, the price-quantity pairs representing the different contracts are equidistant (as in our first numerical study in Table 2). More specifically, the menus with three (five) options have quantity values $\{q^s - 6, q^s, q^s + 6\}$ ($\{q^s - 6, q^s - 3, q^s, q^s + 3, q^s + 6\}$). Each scenario is evaluated by drawing 100 realizations of the vector of all random variables.

		MC		CNF		Percentage of CNF by Contract Option (%)							
# of Contract Options:		Five	Three	Five	Three	Five					Three		
						O1	O2	O3	O4	O5	O1	O3	O5
D. CV = 0.15	Slope 1	606,764	590,203	124	123	13.7	33.9	21.8	27.4	3.2	24.4	61.8	13.8
R. CV= x	Slope 2	555,655	554,445	117	118	12.0	9.4	40.2	35.0	3.4	11.9	55.9	32.2
	Slope 3	549,689	539,072	116	116	5.2	12.1	30.2	36.2	16.4	7.8	54.3	37.9
D. CV = 0.15	Slope 1	659,109	666,461	138	142	10.9	31.9	29.7	23.9	3.6	33.8	47.2	19.0
R. CV = 3x	Slope 2	598,221	592,582	131	132	6.1	9.2	44.3	38.2	2.3	7.6	69.7	22.7
	Slope 3	595,429	585,797	130	129	6.2	6.2	32.3	37.7	17.7	6.2	48.1	45.7
D. CV = 0.3	Slope 1	692,986	705,191	140	140	12.1	31.4	18.6	35.7	2.1	21.4	62.9	15.7
R. CV = x	Slope 2	661,500	679,119	134	134	10.4	9.0	38.8	39.6	2.2	10.4	53.0	36.6
	Slope 3	662,380	641,916	132	132	4.5	11.4	28.0	37.1	18.9	6.8	51.5	41.7
D. CV = 0.3	Slope 1	728,059	Infeas.	157	Infeas.	9.6	29.3	31.8	26.1	3.2	Infeasible		
R.CV = 3x	Slope 2	699,212	727,830	150	151	5.3	8.0	42.7	41.3	2.7	6.6	64.2	29.1
	Slope 3	750,188	721,211	148	148	5.4	5.4	29.7	39.9	19.6	5.4	43.2	51.4

Table 3 Impact of menu design on manufacturer's performance measures

D.CV: Demand Coefficient of Variation, R.CV: Rainfall Coefficient of Variation, MC: Manufacturer Cost, CNF: Contracted Number of Farmers, Oi: Contract Option i

In the first segments of Tables 3 and 4, we report various performance measures under each of the six sets of menus. In Table 3, Columns 3 and 4 specify the expected costs to the manufacturer under each of the six sets of menus. The fifth and sixth columns specify the number of selected farmers, and the remaining columns display what percentage of farmers, out of the set of "selected" farmers, opt for each of the contract options, again under the six sets of menus. (Recall, the options are numbered in increasing order of their quantity volume.)

Across all six sets of menus, we have observed that an expected supply, roughly 16% higher than the expected demand needs to be targeted. While the expected aggregate supply is virtually invariant with respect to the menu choice, Table 3 shows that the number of contracted farmers varies between 116 and 124, i.e., a 7% variation in the supplier base. The expected cost to the

		AFP		AFYM (AFYSD)		AFSM (AFSSD)	
# of contract options:		Five	Three	Five	Three	Five	Three
D. CV = 0.15	Slope 1	2,418	2,365	19.32 (1.59)	19.24 (1.56)	18.22 (1.23)	18.31 (1.21)
R. CV= x	Slope 2	2,335	2,331	19.52 (1.53)	19.56 (1.57)	19.22 (1.39)	19.13 (1.39)
	Slope 3	2,352	2,316	19.68 (1.57)	19.55 (1.57)	19.57 (1.50)	19.40 (1.49)
D. CV = 0.15	Slope 1	2,290	2,265	17.66 (4.37)	17.45 (4.41)	16.58 (3.67)	15.95 (3.52)
R. CV = 3x	Slope 2	2,211	2,185	17.72 (4.41)	17.64 (4.43)	17.30 (4.09)	17.11 (4.04)
	Slope 3	2,209	2,176	17.73 (4.48)	17.68 (4.47)	17.49 (4.29)	17.41 (4.26)
D. CV = 0.3	Slope 1	2,570	2,517	19.23 (1.51)	19.11 (1.49)	18.24 (1.19)	18.23 (1.16)
R. CV = x	Slope 2	2,503	2,493	19.27 (1.45)	19.38 (1.50)	19.00 (1.33)	19.00 (1.33)
	Slope 3	2,524	2,502	19.48 (1.49)	19.42 (1.49)	19.38 (1.43)	19.28 (1.42)
D. CV = 0.3	Slope 1	2,382	Infeas.	17.32 (4.37)	Infeas.	16.36 (3.74)	Infeas.
R.CV = 3x	Slope 2	2,336	2,333	17.43 (4.40)	17.49 (4.40)	17.06 (4.12)	17.03 (4.06)
	Slope 3	2,352	2,337	17.54 (4.44)	17.52 (4.43)	17.33 (4.27)	17.29 (4.25)

Table 4 Impact of menu design on contracted farmers' performance measures

D.CV: Demand Coefficient of Variation, R.CV: Rainfall Coefficient of Variation, AFP: Average Farmer Profit, AFYM: Average of Farmer Yield Means, AFYSD: Average of Farmer Yield Standard Deviations, AFSM: Average of Farmer Supply Means, AFSSD: Average of Farmer Supply Standard Deviations

manufacturer exhibits an even larger variation of 12%. More specifically, the slope of the contract menu line, hence the price sensitivity with respect to quantity variations in the contract, has a significant impact on overall costs. Slope 3 allows for a 9.5% reduction in overall expected costs, compared to Slope 1, assuming the menu consists of five contract choices. An additional 2% cost saving can be achieved by limiting the number of contract options to three.

As we move from Slope 1 to Slope 3, the price sensitivity of quantity variations in the contract diminishes. This induces more and more farmers to select higher quantity contracts, since the per unit price sacrifice in doing so, is progressively smaller. Indeed the contract choice distribution under Slope 1 (2) majorizes that under Slope 2 (3), both when the menus offer three or five contract options. This shift towards the lower unit price / higher volume contracts explains the above cost savings as we move from Slope 1 to Slope 2 and Slope 3. It also explains why the number of selected farmers can be reduced as we move from Slope 1 to Slope 3.

Table 4 displays the implications the menu choice has for the contracted farmers. The third and fourth columns exhibit the average expected profit level per contracted farmer. The fifth and sixth columns show the average expected *yield* per contracted farmer, under the six sets of menus; the numbers within the parentheses represent the average value of the yield standard deviations. The seventh and eighth columns, similarly, display the average mean and standard deviation of the expected *supply* provided per contracted farmer. Recall that the latter is given by the lower of the yield and the contract volume.

Not surprisingly, the average profit per farmer is lower when the number of options in the contract menu is reduced to three, under each of the menu slopes 1–3. More noteworthy is the fact that the profit reduction is less than 2%. The impact on the manufacturer's cost is similarly small, and not always to the manufacturer's benefit. *Thus, little is lost or gained for either party, when offering simple menus with a small number of options.* We reach the same conclusion when designing the contract menus endogenously; see below.

As we move from Slope 1 to Slope 3, the average yield and average supply increases. This increase is enabled by the selected farmers choosing higher quantity contracts, when the associated unit price reduction becomes smaller; see Table 3. It allows the manufacturer to contract with fewer and more conveniently located farmers, maintaining (approximately) the same aggregate expected supply at a significantly lower cost (by 12%). Since the average farmer loses no more than 4% of his profits, a win-win situation can be created by an additional uniform, modest increase of the unit prices under all contracts.

For the same set of problem instances, we have also explored how the different performance measures are impacted by several risk factors. The second segments in Tables 3 and 4 exhibit the performance measures when the standard deviations of the various rainfall distributions are tripled. The third segments exhibit the performance measures when the volatility of the demand distributions at the plants is doubled, and the last segments exhibit these measures when both risk factors are amplified.

Doubling the coefficient of variation of the demand distributions results in an approximately 13% increase in the targeted expected aggregate supply, to continue to meet the coverage constraint. In other words, the safety supply needs to be doubled, with R_2 increasing from 1.16 to 1.31. To enable this, the manufacturer needs to expand the supplier pool by a similar percentage. Focusing on the most efficient set of menus –with Slope 3 and 3-contract options– the increased demand volatility results in an even larger increase (19%) of the manufacturer cost. On the other hand, an increase of the volatility of the rainfall distributions impacts the manufacturer's cost less severely (by about 8.5%), with similarly sized decreases of the farmers' expected profits, yields and supplies. To hold the cost increase to no more than 8.5%, it becomes optimal to expand the contracted farmer pool by some 11%, under the most cost efficient sets of menus –with Slope 3 and 3-contract options. In other words, it is optimal for the manufacturer to face the increased risks, whether supply or demand risks, by diversifying the supply over a significantly expanded supplier pool.

When the volatility of both the rainfall and demand distributions is increased simultaneously, the above effects are compounded. Compared with the base case, the expected minimal cost for the manufacturer, under the set of menus with Slope 3 and 3-contract options, now increases by 34%, and the required number of contracted farmers grows from 116 to 148. Under the set of menus with Slope 1, it becomes necessary to select all but one of the farmers, when the menu offers five contract choices. When it offers only three, the coverage constraint cannot be met even when contracting with *all* of the farmers. When comparing *among* the six sets of menus, we note that the above described patterns for the base case, in the first segment of Table 3, carry over to all subsequent segments.

We have also experimented with contract menus in which the price-quantity pairs lie on a non-linear curve. To this end, we have adapted the three menu lines to parabolae which share the same midpoint and tangent of that midpoint. Table 11 in Appendix G exhibits the contracts under each of the three parabolae, in each of the six states. In Table 12 of Appendix G, we report both the optimality gaps and the computational times using the $MILP$ heuristic under each of the six sets of menus, for the base setting, as well as settings with doubled standard deviations of the demand distributions, tripled standard deviations of the rainfall distributions, or both. The bounds for the optimality gaps remain, invariably excellent, always below 0.5% and, on average, 0.088%. Computational times are comparable to the times we observed under linear contract menus; they are of the same order of magnitude.

8. Conclusions

We have developed a tactical planning model for a manufacturer contracting with part of a general list of potential farmers as suppliers of a given agricultural commodity. The model allows the manufacturer to select an optimal number and set of farmers to contract with, while determining how supplies generated from the contracts are optimally distributed to meet the demands of each of the manufacturer's production facilities. The manufacturer selects a set of farmers to offer a menu of contracts, which is exogenously specified or endogenously determined. The model is a Stackelberg game with asymmetric information, with the manufacturer as the leader and the selected farmers as the followers. A farmer, when selected, chooses the contract which maximizes his expected profit under an optimal irrigation scheme, by solving a two-stage stochastic program, possibly declining all contracts, if a minimum income level cannot be assured. The manufacturer uses the farmers' random supply quantities, under their best response strategies, to determine his farmer selection and end-of-the-season distribution plan, again by solving a stochastic program.

We have shown how both the farmers' and the manufacturer's problem can be solved to optimality and close to optimality, respectively, even when the number of potential suppliers is 1,000 or larger. To this end, we have developed two approximate solution methods and a Lagrangean dual lower bound. We have also shown how our model can be used to gain various managerial insights. Here are our "top ten" insights:

(a) The extremely low average optimality gap of the Greedy-Add heuristic, one of the two algorithmic approaches, indicates that *robust* farmer selections can be made in an expanding environment. If the manufacturer's demand volume grows from one season to the next, so that the optimal number of contracted farmers increases, very little is lost by adding a new set of farmers to the selected set of the prior season. This allows for robust and long-term relationships with the selected farmers, as long as the farmers' characteristics remain stable.

(b) Among many considered farmer characteristics, three stood out as primary selection criteria. More specifically, the likelihood of a given farmer being selected for contracting increases, as (i) the ratio of the expected payment to the farmer and his expected supply volume decreases, (ii) the coefficient of variation of the farmer's supply volume –a measure of his supply volatility– decreases, and (iii) the distance between the farmer and the closest manufacturing plant decreases. These characteristics are identified with logistic regression models.

(c) An individual farmer can, therefore, improve his chance of being selected for contracting by reducing the volatility of his yield. The latter can be accomplished by implementing standardized cultivation methods and by assuring efficient access to a well to supplement rainfall and surface water when needed. The farmer's attractiveness also increases with an increase in his expected yield per acre.

(d) The above findings show that the selection problem involves complex trade-offs among several farmer characteristics. Our Stackelberg game and associated mathematical program efficiently determine a (close-to-) optimal set of farmers among all 2^I subsets of the I potential farmers. In contrast, if we confined ourselves to simple selection rules, e.g., all suppliers with an expected payment per expected ton supplied below a given threshold, the resulting "best" set of farmers will differ from the above (close-to-) optimal set for a substantial percentage of the required number of farmers.

(e) Designing a contract menu around a given central contract option may impact the total expected cost for the manufacturer as well as the number of contracted farmers significantly. We report on instances where rotating a given contract menu line around a fixed central contract option results in a 12% reduction of aggregate costs and a similar reduction of the number of contracted farmers. In some cases, such rotation of the contract menu may imply the difference between feasible and infeasible instances.

(f) From the farmer's perspective, these rotations have a major impact on the percentage of farmers that opt for the higher volume contracts as opposed to the lower volume ones.

(g) Under the same menu change, the expected profit per contracted farmer in these instances changes much more modestly allowing for *win-win opportunities* for the manufacturer and the contracted farmers. At the same time, as mentioned, fewer farmers are selected for contracting.

(h) When constructing the contract menu endogenously, often a small number of contract options suffices, depending on the degree of heterogeneity among the farmer pool. In other words, little is gained by constructing menus with more than three contract options. The same holds under exogenously specified contract menus. Thus, relatively simple menus often suffice.

(i) Folklore predicts that expected system-wide costs increase when demand and yield volumes become more volatile. However, in this setting, an increase in the rainfall volatility may result in a cost reduction for the manufacturer, as increased yield risks induce farmers to opt for higher volume/lower price contracts to meet their minimum profit level constraints.

(j) As correlations among the farmers' supply variables increase, the manufacturer always needs to increase her total expected supply value. Typically, this implies an increase in the number of contracted farmers, as well. Even more dramatically, feasible instances may become infeasible under increased supply correlations. Increased correlations result in significant substitutions in the optimal farmer pool, often replacing more reliable farmers for more conveniently located ones.

Acknowledgments

We thank Hau Lee, the Associate Editor and referees for their guidance and constructive suggestions, and Mustafa Şahin for his help with the computer codes.

References

Bolton, P., M. Dewatripont. 2005. *Contract Theory*. MIT Press, Cambridge, MA.

Boyabatli, O., P. R. Kleindorfer, S. R. Koontz. 2011. Integrating long-term and short-term contracting in beef supply chains. *Management Sci.* **57**(10) 1771–1787.

Chaturvedi, A., V. Martinez de Albeniz. 2011. Optimal procurement design in the presence of supply risk. *Manufacturing Service Oper. Management* **13**(2) 227–243.

Christensen, L. A., D. E. McElyea. 1988. Toward a general method of estimating productivity-soil depth response relationships. *J. Soil Water Conservation* **43**(2) 199–202.

Cornuejols, G., M. L. Fisher, G. L. Nemhauser. 1977. Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms. *Management Sci.* **23**(8).

Dawande, M., S. Gavirneni, M. Mehrotra, V. Mookerjee. 2013. Efficient distribution of water between head-reach and tail-end farms in developing countries. *Manufacturing Service Oper. Management* **15**(2) 221–238.

de Zegher, J. F., D. A. Iancu, H. L. Lee. 2017. Designing contracts and sourcing channels to create shared value. Forthcoming in *Manufacturing Service Oper. Management*.

Devalkar, S. K., R. Anupindi, A. Sinha. 2011. Integrated optimization of procurement, processing and trade of commodities. *Oper. Res.* **59**(6) 1369–1381.

FAO. 2007. Coping with water scarcity. Food and Agriculture Organization of the United Nations. Agriculture and Consumer Protection Department, Rome.

Federgruen, A., N. Yang. 2008. Selecting a portfolio of suppliers under demand and supply risks. *Oper. Res.* **56**(4) 916–936.

Federgruen, A., N. Yang. 2009. Competition under generalized attraction models: Applications to quality competition under yield uncertainty. *Management Sci.* **55**(12) 2028–2043.

Federgruen, A., N. Yang. 2011. Supply chain management under simultaneous supply and demand risks. H. Gurnani, A. Mehrotra, S. Ray, eds., *Supply Chain Disruptions: Theory and Practice of Managing Risk*. Springer-Verlag London Ltd, Chapter 4.

Grosfeld-Nir, A., Y. Gerchak. 2004. Multiple lotsizing in production to order with random yields: Review of recent advances. *Ann. Oper. Res.* **126** 43–69.

Gruley, B., L. Patton. 2014. To stop the coffee apocalypse, starbucks buys a farm. Https://www.bloomberg.com/news/articles/2014-02-13/to-stop-the-coffee-apocalypse-starbucks-buys-a-farm. Accessed in March 8, 2018.

Huh, W. T., S. Athanassoglou, U. Lall. 2012. Contract farming with possible reneging in a developing country: Can it work? *IIMB Management Review* **24** 187–202.

Huh, W. T., U. Lall. 2013. Optimal crop choice, irrigation allocation, and the impact of contract farming. *Production and Operations Management* **22**(5) 1126–1143.

Kulik, A., H. Shachnai, T. Tamir. 2013. Approximations for monotone and non-monotone submodular maximization with knapsack constraints. *Math. Oper. Res.* **38**(4) 729–739.

Lee, J., V. S. Mirrokni, V. Nagarajan, M. Sviridenko. 2009. Non-monotone submodular maximization under matroid and knapsack constraints. *Proceedings of the 41st annual ACM symposium on Theory of computing, STOC '09*. New York, NY. ACM, 323–332.

Liu, S., K. C. So, F. Zhang. 2010. Effect of supply reliability in a retail setting with joint marketing and inventory decisions. *Manufacturing Service Oper. Management* **12**(1) 19–32.

Lovejoy, W. S. 2006. Optimal mechanisms with finite agent types. *Management Sci.* **52**(5) 788–803.

Lutze, H., O. Ozer. 2008. Promised lead-time contracts under asymmetric information. *Oper. Res.* **56**(4) 898–915.

MacDonald, J. M., P. Korb. 2011. Agricultural contracting update: Contracts in 2008. *Economic Information Bulletin Number 72. U.S. Dept. of Agriculture, Econ. Res. Serv.* .

MacDonald, J. M., J. Perry, M. Ahearn, D. Banker, W. Chambers, C. Dimitri, N. Key, K. Nelson, L. Southard. 2004. Contracts, markets, and prices: Organizing the production and use of agricultural commodities. *Agricultural Economic Report, Number 837. U.S. Dept. of Agriculture, Econ. Res. Serv.* .

Mendelson, H., T. Tunca. 2007. Strategic spot trading in supply chains. *Management Sci.* **53**(5) 742–759.

Nemhauser, G. L., L. A. Wolsey. 1978. Best algorithms for approximating the maximum of a submodular set function. *Math. Oper. Res.* **3**(3).

Palma, C. R. 2004. Conjunctive use of surface water and ground water with quality considerations. *Frontiers Water Resource Econom. Natural Resource Management Policy* **29** 225–240.

Pei, P., D. Simchi-Levi, T. I. Tunca. 2011. Sourcing flexibility, spot trading, and procurement contract structure. *Oper. Res.* **59**(3) 578–601.

Rajib, P. 2014. *Commodity Derivatives & Risk Management*. PHI Learnings, Delhi, India.

Rajib, P. 2015. Indian agricultural commodity derivatives market—in conversation with s sivakumar, divisional chief executive, agri business division, itc ltd. *IIMB Management Review* **11**.

Rao, N. H., P. B. S. Sarma, S. Chander. 1990. Optimal multicrop allocation of seasonal and intraseasonal irrigation water. *Water Resources Res.* **26**(4) 551–559.

Rodriguez-Iturbe, I., D. R. Cox, V. Isham. 1987. Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A* **410**(1839) 269–288.

Sango, D., D. Sarupinda, L. Mawire, C. Nyagumbo, S. Gaba, R. Mukombe, A. Mudhefi, D. Maleni, E. Mudyzvivi, B. Nhau. 2016. The contract farming support program. Post-Harvest Survey Report 2011/2012 Season.

Tan, B., N. Çömden. 2012. Agricultural planning of annual plants under demand, maturation, harvest, and yield risk. *European J. of Operational Research* **220** 539–549.

Wang, Y., W. Gilland, B. Tomlin. 2010. Mitigating supply risk: Dual sourcing or process improvement? *Manufacturing Service Oper. Management* **12**(3) 489–510.

Waymire, E., V. K. Gupta. 1981. The mathematical structure of rainfall representations. review of the stochastic rainfall models. *Wat. Resour. Res.* **17** 1261–1272.

Appendix A: Proofs

Proof of Lemma 1:

By assumption, each of the farmer's parameters $\{W_i, \underline{\pi}_i, c_i, \delta_i^L, \delta_i^H\}$ and the distributions of the random variables $\{R_i, C_i\}$ are drawn from finitely many types. Since there are finitely many contracts (L), all random variables $\{S_i\}$ are uniformly bounded by $q_{max} = \max_{\ell=1,\ldots,L} q_\ell$ and their variances are uniformly bounded from below. Finally, the random variable D^{tot} has a finite third moment. Let

$$r_I^3 = \sum_{i=1}^{I} \mathbb{E}\left(|S_i - \mathbb{E}(S_i)|^3\right) + \mathbb{E}\left(|D^{tot} - \mu_{tot}|^3\right) \quad \text{and} \quad s_I^2 = \sum_{i=1}^{I} \mathrm{Var}(S_i)Y_i + \sigma_{tot}^2.$$

Since all random variables $\{S_i\}$ have uniformly bounded supports, $r_I^3 = O(I)$, so that $r_I = O(\sqrt[3]{I})$. Similarly, $s_I^2 = \Omega(I)$, so that $s_I = \Omega(\sqrt{I})$. It follows that the Lyapunov condition, $\lim_{I\to\infty} \frac{r_I}{s_I} = 0$, is satisfied and the Lyapunov Central Limit Theorem applies.

Proof of Lemma 2:

(a) The function $\phi_{i\ell}(\alpha_i, T_i)$ is concave in T_i, since it is the composition of an increasing (piecewise linear) concave function and the concave function $f_i(\alpha_i, \cdot)$. It follows that the maximand in (19b) is concave in T_i and the characterization of $T_{i\ell}^*(\alpha_i, R_i)$ follows immediately.

(b) Note that the interval $[R_i, R_i + C_i]$ is increasing in R_i: Let $R_i^1 < R_i^2$, for any $T_i^1 \in [R_i^1, R_i^1 + C_i]$ and $T_i^2 \in [R_i^2, R_i^2 + C_i]$, $R_i^2 + C_i \geq \max(T_i^1, T_i^2) \geq T_i^2 \geq R_i^2$, i.e., $\max(T_i^1, T_i^2) \in [R_i^2, R_i^2 + C_i]$. Similarly, $\min(T_i^1, T_i^2) \in [R_i^1, R_i^1 + C_i]$. The maximand in (19b) is independent of R_i, and therefore supermodular in (T_i, R_i). It follows from Lemma 2.8.1 in Topkis (1998) that $\arg\max_{R_i \leq T_i \leq R_i + C_i} \{\phi_{i\ell}(\alpha_i, T_i) - \delta_i^T T_i\}$, the set of optimal T_i values in (19b) is increasing in R_i. It follows that $T_{i\ell}^*(\alpha_i, R_i)$, as the smallest element of this set, is increasing in R_i as well.

(c) The proof is analogous to that of part (b). (d) Immediate from part (a).

Proof of Theorem 1:

Proof: With $\mathcal{I}$ be the set of potential farmers, let $\mathcal{I}^*$ the set of selected farmers, and $\mathcal{J}$ the set of manufacturing plants. Then the optimal cost of problem (3), denoted by $g(\mathcal{I}, \mathcal{J})$ for given realizations of supplies $\{S_i\}$ and demands $\{D_j\}$, is supermodular in $(\mathcal{I}\backslash\mathcal{I}^*, \mathcal{J})$ by Theorem 3.4.1(a) of Topkis (1998) and so is $\mathbb{E}(\Psi(Y))$ by the fact that the expectation operator preserves supermodularity. The first term in (1) is a modular set function and, in particular, a supermodular function. Since supermodularity is also closed under summation, the complete objective function (1) is also supermodular as a function of $(\mathcal{I}\backslash\mathcal{I}^*)$.

Appendix B: Lower Bound Computation – Lagrangean Dual

In this appendix, we derive a lower bound for the manufacturer's problem, (M'), which can be computed with modest effort even for large size problem instances, with hundreds or thousands of farmers. The lower bound is obtained by dualizing the constraints (6) and (8) with Lagrange multipliers $\lambda \geq 0$ and $-\infty < \rho < \infty$. We then obtain the following Lagrangean dual of the manufacturer's problem:

$$\max_{\lambda,\rho} \min_{Y,Y_E} \quad L(Y, Y_E, \lambda, \rho) \qquad s.t.\ Y_E \geq \mu_{tot} + z_\epsilon \sigma_{tot}, \qquad \text{where}$$

$$L(Y, Y_E, \lambda, \rho) = \sum_{i=1}^{I} v_i Y_i + \mathbb{E}_{\{S_i, D_j, \gamma_{0j}\}} \Psi(Y) + \lambda \left[-(Y_E - \mu_{tot})^2 + z_\epsilon^2 \sigma_{tot}^2 + z_\epsilon^2 \sum_{i=1}^{I} Var(S_i) Y_i \right] + \rho \left[Y_E - \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i \right]$$

For any fixed pair of Lagrange multipliers λ and ρ, the Lagrangean relaxation decomposes into an optimization problem in the vector of Y and a separate problem in the single continuous variable Y_E. In particular, for fixed λ and ρ, we can write $L(Y, Y_E, \lambda, \rho) = L_1(Y, \lambda, \rho) + L_2(Y_E, \lambda, \rho)$ where

$$L_1(Y, \lambda, \rho) = \sum_{i=1}^{I} v_i Y_i + \mathbb{E}_{\{S_i, D_j, \gamma_{0j}\}} \Psi(Y) + \lambda z_\epsilon^2 \sum_{i=1}^{I} Var(S_i) Y_i - \rho \sum_{i=1}^{I} \mathbb{E}(S_i) Y_i$$
$$L_2(Y_E, \lambda, \rho) = -\lambda (Y_E - \mu_{tot})^2 + \rho Y_E + z_\epsilon^2 \sigma_{tot}^2$$

$L_1(Y, \lambda, \rho)$ represents a supermodular set function, since it equals $g(Y)$ plus two linear functions of Y, see Theorem 1. Hence, $L(Y, Y_E, \lambda, \rho)$ is also a supermodular function of Y.

B.1. Solution of the Lagrangean Relaxation for Fixed (λ, ρ)

We now show how the Lagrangean relaxation can be solved for fixed λ and ρ. We first show how we can compute the optimal value of Y_E. Note that $L_2(Y_E, \lambda, \rho)$ is a concave function of Y_E, so it achieves its minimum in one of the two end points of its feasible interval $[d_0, d_N]$ derived in Section 6.2. (Recall $d_0 = \mu_{tot} + z_\epsilon \sigma_{tot}$ and $d_N = \sum_{i=1}^{I} \mathbb{E}(S_i)$) Therefore, we can find the optimal value of Y_E by comparing $L_2(d_0, \lambda, \rho)$ and $L_2(d_N, \lambda, \rho)$:

$$Y_E^*(\lambda, \rho) = d_0 \iff -\lambda (d_0 - \mu_{tot})^2 + \rho d_0 \le -\lambda (d_N - \mu_{tot})^2 + \rho d_N$$
$$\iff \lambda [d_N^2 - d_0^2 - 2(d_N - d_0)\mu_{tot}] \le \rho (d_N - d_0) \iff [d_N + d_0 - 2\mu_{tot}] \le \frac{\rho}{\lambda}$$

Conversely, $\frac{\rho}{\lambda} < [d_N + d_0 - 2\mu_{tot}]$ if and only if $Y_E^*(\lambda, \rho) = d_N$.

Thus, substituting Y_E by $Y_E^*(\lambda, \rho)$, the Lagrangean relaxation can be written as:

$$z^*(\lambda, \rho) \equiv \min_{Y, \nu} \sum_{k=1}^{K} \sum_{i=0}^{I} \sum_{j=1}^{J} \delta_k \gamma_{ij} \nu_{ijk} + \sum_{i=1}^{I} [v_i - \rho \mathbb{E}(S_i) + \lambda z_\epsilon^2 Var(S_i)] Y_i + \lambda [-(Y_E^* - \mu_{tot})^2 + z_\epsilon^2 \sigma_{tot}^2] + \rho Y_E^* \tag{31}$$

$$s.t. \sum_{j=1}^{J} \nu_{ijk} \le S_{ik} Y_i \quad i = 0 \dots I \quad k = 1 \dots K$$
$$\sum_{i=0}^{I} \nu_{ijk} \ge D_{jk} \quad j = 1 \dots J \quad k = 1 \dots K; \quad \nu_{ijk} \ge 0, \; Y_i \in \{0, 1\}$$

which is linear in the decision variables Y and ν; hence it is a $MILP$.

B.2. Solution Methods for $z_D = \max_{\lambda, \rho} \; z^*(\lambda, \rho)$

It is well known and easily verified that the function $z^*(\lambda, \rho)$ is jointly concave in (λ, ρ); however, as the optimal value of a mixed integer program, it fails to be differentiable everywhere. Nevertheless, a steepest ascent subgradient method can be employed to find the maximizing pair of Lagrange multipliers (λ^*, ρ^*). However, to ensure convergence, very small step sizes need to be chosen in the steepest ascent method.

We have observed that the following tâtonnement scheme converges considerably faster: at odd (even) numbered iterations, the scheme fixes the last obtained value of ρ (λ) and finds the corresponding optimal value of λ (ρ) via a standard bisection method. (In view of the joint concavity of the function $z^*(\lambda, \rho)$, the tâtonnement scheme is guaranteed to converge to an optimum solution.)

Solving the Lagrangean dual z_D generates a very useful lower bound against which various heuristics may be benchmarked. In addition, the optimal solution vector $Y^*(\lambda^*, \rho^*)$ in the Lagrangean relaxation (31) provides another promising initial set of farmers to start the *Greedy-add* heuristic with.

Appendix C: Analytical Expressions for the Coefficients in (M^{cont})

In this appendix, develop analytical expressions for the coefficients in the mathematical program (M^{cont}) used to determine optimal contract menus. For notational simplicity, we assume, as in Lemma 2(d), that the yield functions $f_i(\alpha_i, T_i)$ are increasing and differentiable, so that the analytical characterization of the optimal aggregate water supply $T^*_{i\ell}(\alpha_i, R_i)$ provided there, can be employed. Since, under *all* yield functions, $T^*_{i\ell}(\alpha_i, R_i)$ is an increasing function of R_i, see Lemma 2(b), define $R^*_i(\alpha_i, q_\ell)$ as the unique root of the equation $T^*_{i\ell}(\alpha_i, R_i) = t^*_i(\alpha_i, q_\ell)$, i.e., $R^*_i(\alpha_i, q_\ell)$ is the inverse function ${T^*_{i\ell}}^{-1}(\alpha_i, \cdot)$ evaluated at $t^*_i(\alpha_i, q_\ell)$ and denotes the break-even rainfall quantity under which the optimal total water supply suffices to yield the quantity value of contract ℓ.

$$\begin{aligned}
U_i(\ell|k, q_\ell) =& \delta^T_i \mathbb{E}(R_i) + (P(q_\ell) - c_i) \int_0^{R^*_i(\alpha^k, q_\ell)} f_i(\alpha^k, T^*_{i\ell}(\alpha^k, R_i)) dG_i(R_i) + (P(q_\ell) - c_i) q_\ell \overline{G}_i(R^*_i(\alpha^k, q_\ell)) \\
&+ (\underline{p}^s_i - c_i) \int_{R^*_i(\alpha^k, q_\ell)}^{\infty} \left[f_i(\alpha^k, T^*_{i\ell}(\alpha^k, R_i)) - q_\ell \right] dG_i(R_i) - \delta^T_i \int_0^{\infty} T^*_{i\ell}(\alpha^k, R_i) dG_i(R_i) \qquad (32)
\end{aligned}$$

$$\mathbb{E}(S_i|q) = \sum_{k=1}^{K} \mathbb{P}[\alpha_i = \alpha^k] \left\{ \int_0^{R^*_i(\alpha^k, q_k)} f_i(\alpha^k, T^*_{ik}(\alpha^k, R_i)) dG_i(R_i) + q_k \overline{G}_i(R^*_i(\alpha^k, q_k)) \right\} \qquad (33)$$

$$\mathbb{E}(S^2_i|q) = \sum_{k=1}^{K} \mathbb{P}[\alpha_i = \alpha^k] \left\{ \int_0^{R^*_i(\alpha^k, q_k)} f^2_i(\alpha^k, T^*_{ik}(\alpha^k, R_i)) dG_i(R_i) + q^2_k \overline{G}_i(R^*_i(\alpha^k, q_k)) \right\} \qquad (34)$$

$$Var(S_i|q) = \mathbb{E}(S^2_i|q) - \mathbb{E}^2(S_i|q) \qquad (35)$$

$$v_i(q) = \sum_{k=1}^{K} \mathbb{P}[\alpha_i = \alpha^k] P(q_k) \left\{ \int_0^{R^*_i(\alpha^k, q_k)} f_i(\alpha^k, T^*_{ik}(\alpha^k, R_i)) dG_i(R_i) + q_k \overline{G}_i(R^*_i(\alpha^k, q_k)) \right\} \qquad (36)$$

Appendix D: Numerical Analysis: Scenario Generation

The potato chips manufacturer contracts with several thousands of farmers located in 19 districts across six states of India. (India is divided into 36 states or territories each of which is subdivided into districts.)

Indian farmers are typically very small. Table 1 in Singh et al. (2002) shows that in 1991, country-wide, a total of 165.6 million hectares were cropped, by no less than 106.6 million farmers. (The *total* cropped area barely grew between 1971 and 1991; however, the number of farmers grew from 70.5 million in 1971 to 106.6 million in 1991. This, of course, implies that the average farm size decreased significantly in this time span.) The table partitions the farms into five categories, according to the size of the cropped area: sub-marginal, marginal, small, medium, and large farms.

We only had access to aggregate data for the 19 districts, rather than data for each of the contracted or potential farmers. When generating our problem instances, we assumed that the manufacturer only considered small, medium, and large farms, which, nationwide, represent 41% of all farms. We assumed that the distribution of farm sizes among the list of potential farms, in each district, reflected the national distribution as in Table 1. For any given farmer, in any of the districts, we therefore assigned a land size that is drawn, independently, from this national three-point distribution.

D.1. Geographic distribution of the farmers

We were given data for the *total contracted* cropped area in each district, for the years 2009-2010. For any problem instance with I country-wide potential farmers, we distributed them across the 19 districts in proportion to the districts' average cropped area in the above two calendar years. (The resulting district numbers were adjusted by a factor uniformly chosen from the interval $[0.9, 1.1]$ and then rounded to the nearest integer.) Table 10 in Appendix G lists the 19 districts and the percentage of the total cropped area in each district. In view of the above distribution procedure, we have that, in each of the country-wide instances in Section 7, the percentages listed in Table 10 approximately mirror the distribution of the total of I farmers among the districts.

D.2. The farmer's supply quantity distribution

In terms of the yield functions, $f_i(\alpha_i, T_i)$, we chose *structure 1* see (10), specified as follows:

$$X_i = \Lambda_i \sum_{s=1}^{6} \alpha_s I_{is} - \beta[W_i - R_i - x_i]^+, \text{ where} \tag{37}$$

$$I_{is} = \begin{cases} 1, & \text{if farmer } i \text{ is located in state } s,\ s = 1, \dots, 6;\ i = 1, \dots, I \\ 0, & \text{otherwise,} \end{cases}$$

Λ_i is the land size of farmer i (in hectares), $\beta > 0$ is a non-negative parameter, and $\{\alpha_1, \dots, \alpha_6\}$ are random variables. Thanks to data provided by the Columbia Water Center, we had access to a 100-year time series, for each of the 19 districts, of the *total* rainfall (in millimeters –mm) over the course of the 90 day growing season. We assumed that each farmer experiences a random rainfall quantity generated from a Normal distribution whose mean and standard deviation match the 100-year distribution pertaining to the district the farmer resides in. The Center also provided an estimate of the *ideal* water supply WS_d (in mm) for each district $d = 1, \dots, 19$. However, this number represents the ideal water supply, assuming *constant* daily supplies. In practice, rain falls periodically, with many dry days in between rainy episodes, generating major deficits even when the aggregate season-wide rainfall is well in excess of the ideal aggregate water supply volume. (Recall the alternative deficit specification in equation (14) based on the daily rainfall process.) Since we had no access to daily rainfall data, we scaled the numbers $\{WS_d,\ d = 1, \dots, 19\}$ up by a factor $f = 1.2/(\frac{1}{19}\sum_{d=1}^{19} WS_d/\eta_d)$ with η_d, the mean season-wide rainfall in district d, computed from the above 100-year distribution, resulting in adjusted numbers $\{\widehat{WS}_d,\ d = 1, \dots, 19\}$. In each problem instance, we specified each farmer's well capacity as 25% of his rainfall quantity.

Since $W_i = \Lambda_i \widehat{WS}_{d(i)}$ and $R_i = \Lambda_i \rho_{d(i)}$, with $d(i)$ the district to which farmer i belongs, we get the following final specification of the yield equation (37):

$$X_i = \Lambda_i \sum_{s=1}^{6} \alpha_s I_{is} - \beta[\Lambda_i \widehat{WS}_{d(i)} - \Lambda_i \rho_{d(i)} - x_i]^+ \tag{38}$$

We estimated the distributions of the six state specific intercepts $\{\alpha_s : s = 1, \dots, 6\}$ and the value of the slope β by running a regression based on the 38 observations of supply volumes per district, for a unit (1 hectare) of land, in the two 2009 and 2010 contract years. (When calculating the deficit quantities in

(38), we assumed no water was drawn from wells, i.e., $x_i = 0$.) The regression equation had an adjusted R^2 value of 0.92. The point estimates and standard errors for the intercept values are reported in Table 5; they show significant differences in the farm productivity across the six states, with Uttar Pradesh roughly 2.5 times as productive as the state of Karnataka. In our problem instances, we assigned to each farmer i, the yield function $f_i(\cdot,\cdot)$ in (38), with an α-value drawn from a 5-point distribution, anchored on a Normal distribution with mean and standard deviation given by the estimated mean μ_s and standard error σ_s of the $\alpha-$parameter pertaining to the state in which the farmer resides. (The 5-point distribution used the values $\{\mu_s - 2\sigma_s,\ \mu_s - \sigma_s,\ \mu_s,\ \mu_s + \sigma_s,\ \mu_s + 2\sigma_s\}$ as its support, with probabilities $\{\Phi^{-1}(-1.5),\ \Phi^{-1}(-0.5) - \Phi^{-1}(-1.5),\ \Phi^{-1}(0.5) - \Phi^{-1}(-0.5),\ \Phi^{-1}(1.5) - \Phi^{-1}(0.5),\ 1 - \Phi^{-1}(1.5)\}$ to match the underlying distribution; a discrete distribution was used to be consistent with our model assumption of finitely many farmer types.)

State	Mean	Standard Error	State	Mean	Standard Error
West Bengal	18.02	1.16	Uttar Pradesh	29.36	3.82
Bihar	18.30	1.99	Maharashtra	12.62	1.59
Gujarat	14.85	3.32	Karnataka	11.49	1.29

Table 5 Estimations of the parameters in equation (38)

In this numerical study, we assumed that the farmer's intercept value α is his private information while the manufacturer only knows the distribution from which it is drawn. No other sources of asymmetric information are assumed. For the slope β, we uniformly used the number $\beta = 0.0009$, which corresponds to the lower point of the 95% confidence interval of the estimate of this parameter.

D.3. The farmer's cost parameters and menu of contracts

We obtained state-specific values for the cost parameters c_i, the variable cost of producing a metric ton (MT) of potatoes, exclusive of irrigation costs, and the irrigation cost per cubic meter of water. The former cost value includes the cost of seeds, planting, ploughing, fertilizers, plant protection chemicals, harvesting, and packing costs.

We specified a menu of contracts, differentiated by state and farm size category. (This represents a total of 18 menus.) Each menu is characterized by a *linear* price-quantity relationship. The menus were constructed as follows: For each state s, we were given the average price, p^s, paid per MT of potatoes, as well as the average quantity, q^s, procured per hectare. As an example, consider the contract menu for small farmers: The point $(p^s,\ 1.43q^s)$ was selected as the *midpoint* on the menu, since an average small farmer has a land size of 1.43 hectares, see Table 1. When quantities are displayed on the $X-$axis and prices on the vertical $Y-$axis, the line with slope $-3/2$ was drawn through this point, and 5 equidistant points were selected on either side of the midpoint. For the "medium" and "large" farmers, the same procedure was used except that the midpoint was selected as $(p^s,\ 2.76q^s)$ and $(p^s,\ 7.95q^s)$, reflecting their respective land sizes, see Table 1, and the slope of the contract line as -1 and $-3/5$, respectively.

D.4. The manufacturer's demands and transportation costs

The manufacturer has three production facilities. In our study, we drew the product demands in these facilities from a Normal distribution with a coefficient of variation of 0.15. For each problem instance, we generated a "high demand version" and "low demand version" in which the expected aggregate demand is specified as 65%

and 50% of the aggregate supply across all I potential suppliers, respectively, assuming each farmer supplies the quantity corresponding with the midpoint of his contract line. Finally, we generated transportation cost rates $\{\gamma_{ij}\}$ by multiplying a given average cost rate per MT per mile, with the relevant, Google Maps calculated, road distance between the centroid of the farmer's district and the relevant production facility.

We used the following specification of the $\{\gamma_{0j}\}$ coefficients.

$$\gamma_{0j} = 10 * [\max_{i,j} \gamma_{ij}] + \mathbb{E}(p_s), \qquad j = 1, \dots, 3 \tag{39}$$

The first term represents an upper bound for the per unit distribution cost. When modeling an explicit linkage between the $\{\gamma_{0j}\}$-value and the aggregate rainfall quantity, (39) was modified to

$$\gamma_{0j} = 10 * [\max_{i,j} \gamma_{ij}] + \mathbb{E}(p_s) - 0.084 * STD(p_s) * \left[\frac{R^{tot} - \mathbb{E}(R^{tot})}{STD(R^{tot})} \right], \tag{40}$$

adopting the estimated correlation coefficient in Bhanumurthy et al. (2013).

When evaluating the set function $g(Y)$ in the manufacturer's problem, see (1), we evaluated the second term in (1) by drawing a sample of 100 $\{S_i,\ D_j :\ i = 1, \dots, I\ j = 1, \dots, J\}$ vectors. To solve each of the farmer problems, we drew a sample of 10,000 realizations for the vectors $\{R_i, C_i\}$, when evaluating each of the possible contracts in the offered menu, see problem (F) and equation (15).

All reported computation times refer to a laptop with Intel(R) Core(TM) i7-3537U CPU @ 2.00 GHz and 2.50 GHz processor, 8GB RAM and 64-bit Windows operating system.

Appendix E: A Logit Model to Identify Farmer Attributes as Selection Criteria

The following logit models provided the best fit to explain which farmers are selected as a function of several attributes. Let

$$Y_0 = \begin{cases} 1, & \text{if } t_i + \epsilon_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

with $t_i = \beta_0 + \beta^T X_i$, where X_i is a vector of farmer characteristics, and ϵ_i an unobserved error term. The following models provided the best overall fit:

$$t_i = \underset{(0.01)}{117.6} - \underset{(0.01)}{0.36} (v_i/\mathbb{E}(S_i)) - \underset{(0.56)}{11.86}\ CV(S_i) - \underset{(0.0006)}{0.36}\ \Delta_i \qquad \text{instance in } * \tag{41}$$

$$t_i = \underset{(<0.0001)}{23.2} - \underset{(<0.0001)}{0.01} (v_i/\mathbb{E}(S_i)) - \underset{(<0.0001)}{22.9}\ CV(S_i) - \underset{(0.1)}{0.03} \Delta_i \qquad \text{instance in } ** \tag{42}$$

where $(v_i/\mathbb{E}(S_i))$ denotes the expected payment to the farmer divided by his expected supply volume, $CV(S_i)$ denotes the coefficient of variation of his supply volume, and Δ_i denotes his distance to the nearest manufacturing plant. The numbers within parentheses represent the logistic regression's equivalent of the p-values in ordinary least-squares regression. Beyond the generally low "p"-values, the goodness-of-fit in (41) and (42) is further demonstrated by a percent concordants of 92% and 90%, respectively.

Appendix F: Numerical Analysis with Endogenously Determined Contract Menus

In parallel to the numerical study reported in "Impact of the Menu Design" part of Section 7, we have conducted a parallel study in which the contract menu is designed *endogenously*, based on the methodology of Section 5. More specifically, based on a set of $I = 157$ potential farmers and their characteristics, we have evaluated what *optimal* price-quantity pairs should be offered in each of the 18 districts considered in Section

7 on the contract line with Slope 2. (The above farmer distribution procedure ended up generating farmers in only 18 of the 19 districts.) In other words, instead of using five arbitrary and exogenously given contracts on the menu line, we determine the optimal combination of contracts by solving the mathematical program (M^{cont}), for each district separately. (For the purpose of this experiment, we assigned each district a fraction of the countrywide (random) demand volume in proportion to the district's total expected yield among all farmers.) For this experiment, we also incorporated the spot market option for the farmers' excess yields by choosing a state specific (lower bound on) spot market price, $\underline{p}_i^s$, in the farmer's problem (F) as follows: $\underline{p}_i^s = \max\{0.8 * \min[\text{average spot price in 2009, average spot price in 2010}], 1.1c_i\}$; see Table 6 for the spot price data. In terms of the individual rationality constraints (IR) in (27), we selected the minimum expected profit levels for each of the five types as follows: starting with the type-1 farmers (with the stochastically smallest yield distributions), we selected a minimum profit level below the maximum achievable profit (under any contract on the menu line). After specifying the minimum profit level for type i, $i = 1, \ldots, 4$, we selected a minimum profit level for type $i + 1$, above that of type i, but below its maximum achievable profit level, see Figure 1 in Appendix G. Table 7 displays the minimum profit levels and optimal contract quantities for all five types in each of the 18 districts. In many cases the optimal solution uses less than five distinct points; the average optimal number of contracts is 3.17.

		avg. spot price					avg. spot price		
S	State	2009 ($/MT)	2010 ($/MT)	$\underline{p}_i^s$ ($/MT)	S	State	2009 ($/MT)	2010 ($/MT)	$\underline{p}_i^s$ ($/MT)
1	West-Bengal	236*	104	84	4	Uttar-Pradesh	177	93	131
2	Bihar	234	150	120	5	Maharashtra	258	154	123
3	Gujarat	145*	160*	130	6	Karnataka	258	216	173

Table 6 State summary statistics. S: state index. *: Based on incomplete spot price data

Finally, we have explored how increased volatility of the rainfall and demand distributions impacts the optimal menu choices and associated expected manufacturer's cost. We report the results in Table 8 for District 1 and District 16, when the coefficient of variation of all demand and rainfall distributions are increased by 50% and 250%. For District 1, we observe an interesting phenomenon: the increased volatility benefits the manufacturer resulting in a (modest) cost saving, because majority of the farmers can now be induced to accept higher quantity values, and hence lower prices (58% of the farmers are of types 2 and 3). The reason for this upward shift of the contract volume is that, in order to continue to meet the minimum expected profit level, these farmers need a higher contract volume to take advantage of better than average yields. Increased rainfall volatilities do result in the farmers being worse off, but the average profit reduction is strongly mitigated by the upward adjustments of the contract quantities. Increased demand volatility results in higher costs for the manufacturer. When the volatility of both the demand and rainfall distributions are amplified there is no feasible contract menu, given the coverage and all minimum expected profit level constraints.

		$\underline{\pi}_{ik}$: min. acceptable profits					optimal quantities (prices) in the menu				
S	D	$\underline{\pi}_{i1}$	$\underline{\pi}_{i2}$	$\underline{\pi}_{i3}$	$\underline{\pi}_{i4}$	$\underline{\pi}_{i5}$	q_1^*	q_2^*	q_3^*	q_4^*	q_5^*
1	1	1700	1750	1800	1850	1900	23.13(151.1)	23.88(149.9)	23.88(149.9)	26.70(145.2)	29.44(140.6)
1	2	1700	1750	1800	1850	1900	21.75(153.4)	24.00(149.7)	24.00(149.7)	28.01(143.0)	28.01(143.0)
1	3	1700	1750	1800	1850	1900	22.34(152.4)	22.54(152.1)	25.19(147.7)	25.19(147.7)	28.30(142.5)
1	4	1700	1750	1800	1850	1900	21.42(154.0)	23.72(150.1)	23.72(150.1)	27.06(144.6)	27.06(144.6)
1	5	1700	1750	1800	1850	1900	23.37(150.7)	24.08(149.5)	24.13(149.5)	27.01(144.6)	27.01(144.6)
1	6	1700	1750	1800	1850	1900	22.03(152.9)	22.74(151.8)	25.45(147.2)	25.45(147.2)	28.06(142.9)
2	7	1600	1700	1850	2000	2100	19.11(157.8)	19.11(157.8)	19.11(157.8)	19.11(157.8)	19.11(157.8)
3	9	75	300	600	800	950	5.85(197.3)	12.80(185.7)	12.80(185.7)	12.80(185.7)	17.76(177.4)
3	10	200	400	600	800	1000	18.54(176.1)	18.54(176.1)	18.54(176.1)	18.54(176.1)	18.54(176.1)
4	11	1050	1150	1200	1250	1350	22.71(172.5)	23.54(171.1)	24.06(170.2)	24.06(170.2)	24.06(170.2)
4	12	1100	1200	1250	1300	1350	20.56(176.1)	26.38(166.4)	26.38(166.4)	26.38(166.4)	26.38(166.4)
5	13	2900	3325	3750	4150	4550	13.35(323.1)	15.55(319.4)	17.86(315.6)	21.29(309.9)	22.49(307.9)
5	14	2800	3225	3650	4050	4425	13.32(323.1)	15.53(319.4)	18.92(313.8)	20.13(311.8)	22.34(308.1)
5	15	2050	2500	2950	3350	3750	11.85(325.6)	16.74(317.4)	16.74(317.4)	18.53(314.5)	20.89(310.5)
6	16	1350	1500	1650	1750	1900	13.66(210.9)	14.16(210.1)	14.81(209.0)	14.81(209.0)	14.81(209.0)
6	17	1300	1450	1600	1750	1850	13.98(210.4)	13.98(210.4)	14.78(209.0)	14.78(209.0)	14.78(209.0)
6	18	1200	1350	1500	1650	1750	10.75(215.8)	13.79(210.7)	13.79(210.7)	13.79(210.7)	13.79(210.7)
6	19	1350	1500	1640	1750	1900	13.67(210.9)	14.23(210.0)	14.78(209.0)	14.78(209.0)	14.78(209.0)

Table 7 Minimally acceptable profits and the associated optimal contract quantities for five types of farmers in each district. Numbers in parentheses are the associated price values. S: state index, D: district index.

District 1						
	opt. quantities (prices) in the menu					
	q_1^*	q_2^*	q_3^*	q_4^*	q_5^*	Manuf. Cost
D.CV = x, R.CV = y	23.13 (151.1)	23.88 (149.9)	23.88 (149.9)	26.7 (145.2)	29.44 (140.6)	44,003.5
D.CV = x, R.CV = 1.5y	22.13 (152.8)	24.3 (149.2)	24.3 (149.2)	26.3 (145.8)	28.53 (142.1)	43,837.3
D.CV = 3.5x, R.CV = y	21.58 (153.7)	25.04 (147.9)	25.04 (147.9)	26.83 (144.9)	29.44 (140.6)	44,505.4
D.CV = 3.5x, R.CV = 1.5y	Infeasible					N/A
District 16						
	opt. quantities (prices) in the menu					
	q_1^*	q_2^*	q_3^*	q_4^*	q_5^*	Manuf. Cost
D.CV = x, R.CV = y	13.66 (210.9)	14.16 (210.1)	14.81 (209)	14.81 (209)	14.81 (209)	12,157.0
D.CV = x, R.CV = 1.5y	13.34 (211.4)	13.49 (211.2)	15.56 (207.7)	15.56 (207.7)	15.56 (207.7)	12,369.6
D.CV = 3.5x, R.CV = y	13.66 (210.9)	15.11 (208.5)	16.82 (205.6)	18.14 (203.4)	18.14 (203.4)	13,387.6
D.CV = 3.5x, R.CV = 1.5y	Infeasible					N/A

Table 8 Impact of risk measures on manufacturer performance measures. Numbers in parentheses are the associated price values. D.CV: Demand Coefficient of Variation, R.CV: Rainfall Coefficient of Variation

Appendix G: Tables and Figures

Commodity	91-93	96-97	01-02	05	08	Commodity	91-93	96-97	01-02	05	08
All commodities	28.8	32.1	37.7	40.7	38.5						
Crops	24.7	22.8	27.8	29.9	27.3	Peanuts	47.5	34.4	28	65.3	73.1
Corn	11.3	12.9	14.8	19.6	26.1	Tobacco	0.3	0.3	52.7	79.3	99.3
Soybeans	10.1	13.4	9.4	18.4	25.1	Cotton	30.4	33.8	52.6	45	36.2
Wheat	5.9	9	6.5	7.5	22.5	Fruit	na	41.7	41.9	48.9	38.4
Sugar Beets	91.1	75.2	96.7	82.1	90.8	Vegetables	na	28	28.2	40.9	39.3
Rice	19.7	25.9	38.7	27.1	45.4	Other	7.8	23.8	39.5	25.9	22.5

Table 9 Share of commodity production under contract (%), by commodity. na = data not available.

S	State	D	District	%CA (%)	$\mathbb{E}(R_d)$ (m^3/ha)	STD(R_d) (m^3/ha)	WS_d (m^3/ha)
1	West-Bengal	1	Bankura	7.40	11,070	1,832	11,078
		2	Burdwan	4.83	10,755	1,886	11,262
		3	Hooghly	3.66	11,538	2,073	9,562
		4	Howrah	2.28	12,002	2,342	9,075
		5	Midnapore	1.50	11,899	2,023	4,978
		6	Birbhum	0.80	11,275	2,140	11,318
2	Bihar	7	Katihar	0.95	11,464	2,650	11,598
		8	Purnia	0.32	11,570	2,646	12,679
3	Gujarat	9	Dessa	4.72	5,680	2,322	15,753
		10	Sabar Katha	2.54	7,069	2,610	14,767
4	Uttar-Pradesh	11	Agra	2.65	5,911	1,662	15,030
		12	Kannuaj	1.89	6,765	2,044	14,002
5	Maharashtra	13	Satara	36.32	13,074	4,139	8,382
		14	Pune	18.16	11,007	4,705	8,875
		15	Ahmednagar	2.27	5,726	2,385	10,000
6	Karnataka	16	Chikmanglur	2.43	16,564	4,015	7,657
		17	Belgaum	2.43	7,923	2,141	8,057
		18	Dharwad	2.43	6,979	2,957	8,415
		19	Kodagu	2.43	20,526	4,489	5,825

Table 10 District summary statistics. S: State Index, D: District index, %CA: Percentage Cropped Area

Appendix H: Developing Future Spot Price Distributions

The (random) cost coefficients $\{\gamma_{0j},\ j = 1, \ldots, J\}$ used in the second stage of the manufacturer's problem (M) depend on the future spot price at the end of the growing season, say 9 months hence. There are several approaches to develop this distribution.

First, point estimates of the future spot price may be gathered from a group of commodity experts, and the Delphi method used to form a distribution around the collection of point estimates. Second, the current future prices. corrected for capital carrying cost, may be used as the *mean* of the future spot price distribution. Such commodity futures are traded on various exchanges throughout the world, in particular the Kansas City Board of Trade, operating in conjunction with the Chicago Merchandise Exchange and the commodity exchanges in Hannover, Germany and Amsterdam, the Netherlands, and the Security Exchange Board of India, see also Rajib (2014, 2015). Comparing a historical time series of futures' prices with realized

							Linear					Parabola				
qty:	q_1	q_2	q_3	q_4	q_5	prices:	p_1	p_2	p_3	p_4	p_5	p_1	p_2	p_3	p_4	p_5
State 1	19	22	25	28	31	Slope 1	168	158	148	138	128	158	155.50	148	135.50	118
						Slope 2	158	153	148	143	138	153	151.75	148	141.75	133
						Slope 3	152	150	148	146	144	150	149.50	148	145.50	142
State 2	19	22	25	28	31	Slope 1	168	158	148	138	128	158	155.50	148	135.50	118
						Slope 2	158	153	148	143	138	153	151.75	148	141.75	133
						Slope 3	152	150	148	146	144	150	149.50	148	145.50	142
State 3	12	15	18	21	24	Slope 1	197	187	177	167	157	187	184.50	177	164.50	147
						Slope 2	187	182	177	172	167	182	180.75	177	170.75	162
						Slope 3	181	179	177	175	173	179	178.50	177	174.50	171
State 4	32	35	38	41	44	Slope 1	167	157	147	137	127	157	154.50	147	134.50	117
						Slope 2	157	152	147	142	137	152	150.75	147	140.75	132
						Slope 3	151	149	147	145	143	149	148.50	147	144.50	141
State 5	11	14	17	20	23	Slope 1	337	327	317	307	297	327	324.50	317	304.50	287
						Slope 2	327	322	317	312	307	322	320.75	317	310.75	302
						Slope 3	321	319	317	315	313	319	318.50	317	314.50	311
State 6	10	13	16	19	22	Slope 1	227	217	207	197	187	217	214.50	207	194.50	177
						Slope 2	217	212	207	202	197	212	210.75	207	200.75	192
						Slope 3	211	209	207	205	203	209	208.50	207	204.50	201

Table 11 **Contract quantity and price levels for linear and parabola menu lines**

		M's Problem *MILP* Heuristic OG				M's Problem Runtimes (sec)			
	menu line:	Linear		Parabola		Linear		Parabola	
	# of contract options:	Five	Three	Five	Three	Five	Three	Five	Three
demand CV = 0.15	Slope 1	0.050%	0.035%	0.033%	0.043%	30.87	17.49	708.33	16.92
rainfall CV = x	Slope 2	0.002%	0.047%	0.036%	0.024%	7.08	14.49	10.55	17.95
	Slope 3	0.049%	0.048%	0.036%	0.351%	21.91	36.75	58.44	17.86
demand CV = 0.15	Slope 1	0.045%	0.039%	0.044%	0.015%	40.33	21.50	60.84	17.11
rainfall CV = 3x	Slope 2	0.288%	0.050%	0.253%	0.012%	40.87	193.76	22.65	2.52
	Slope 3	0.472%	0.049%	0.453%	0.032%	2.17	31.14	16.01	21.26
demand CV = 0.3	Slope 1	0.050%	0.016%	0.020%	0.012%	24.59	6.85	160.32	7.29
rainfall CV = x	Slope 2	0.024%	0.007%	0.005%	0.025%	8.23	11.81	17.05	87.21
	Slope 3	0.043%	0.011%	0.362%	0.258%	20.31	12.14	4.32	9.78
demand CV = 0.3	Slope 1	0.034%	Infeas.	0.024%	0.004%	3.62	Infeas.	7.91	3.28
rainfall CV = 3x	Slope 2	0.036%	0.050%	0.038%	0.007%	13.85	15.00	9.76	4.94
	Slope 3	0.050%	0.025%	0.014%	0.022%	176.30	6.50	13.64	19.60

Table 12 **Comparison of menus designed based on linear and parabola lines under various risk factors**
OG: Optimality Gap

spot prices (at the corresponding future dates) enables one to calculate the average prediction error and use this measure as the *standard deviation* of the future spot price distribution. The future spot price can then be modeled as a Normal distribution with the above mean and standard deviation.

Finally, for several agricultural commodities, there are liquid markets with option contracts for call or put options on the targeted future dates with different strike prices. Using the prevailing option prices for the different strike values, one may "backout" what the market's anticipated distribution of the future spot price looks like. This analysis is most easily performed by European options, but can be done for American options as well. See, for example, Hull (2013).

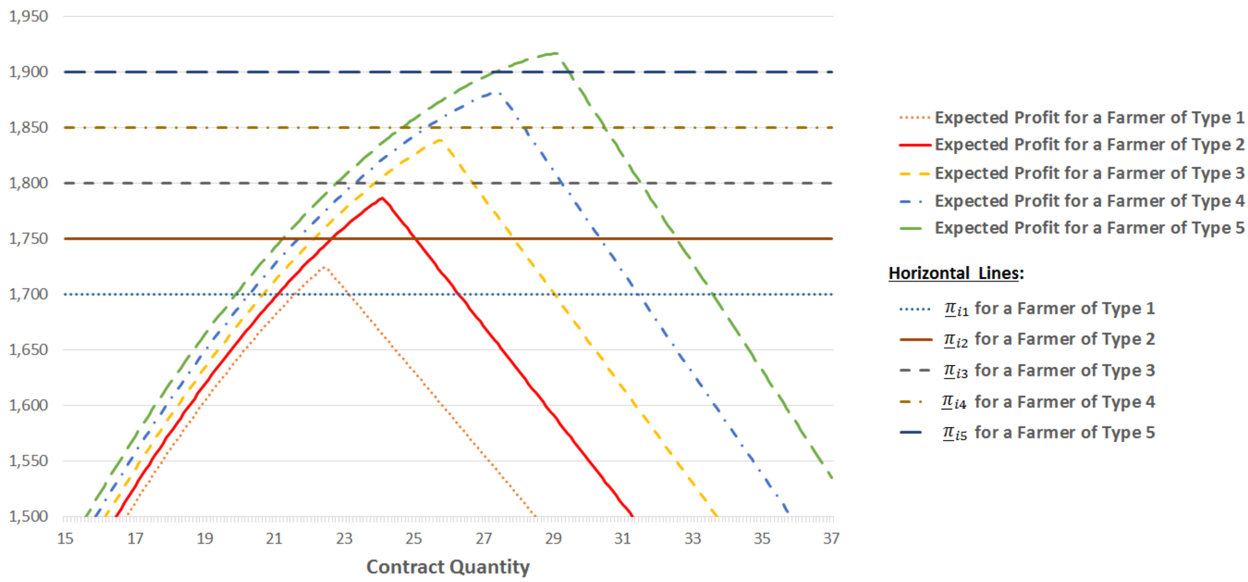

Figure 1 **Expected farmer profit as a function of contract quantity and min. acceptable profits in District 1**

Appendix I: Perfect Information Case

Given the choice of very high cost parameters $\{\gamma_{0j},\ j=1,\dots,J\}$ for procurement from an outside emergency source, it is, under perfect information, optimal, in any scenario, to choose a set of farmers whose aggregate supply covers the aggregate demand, assuming such a set exists. This implies that in the perfect information case, the coverage constraint (2) fails to be relevant, and can be omitted.

The formulation of the perfect information version of the problem requires the following notation: Assume there are K possible scenarios for the realizations of the random variables (i.e., assume all random variables have finite support). Let

- S_{ik} = the supply volume provided by farmer i, under scenario k, as a best response to the offered menu of contracts, $i=1,\dots,I$ and $k=1,\dots,K$
- D_{jk} = the demand volume at the j^{th} manufacturing plant, under scenario k, $j=1,\dots,J$ and $k=1,\dots,K$
- $Y_{ik}=1$ if the manufacturer selects farmer i, under scenario k, $i=1,\dots,I$, $k=1,\dots,K$; and 0, otherwise.
- ν_{ijk} = the number of units shipped from farmer i to manufacturing plant j, under scenario k, $i=1,\dots,I$, $j=1,\dots,J$, and $k=1,\dots,K$
- δ_k: the likelihood associated with scenario k, $k=1,\dots,K$

The "perfect information" version of (M) may then be formulated as:

$$(PIM)\quad \min_{Y,\nu} \sum_{k=1}^{K}\delta_k\left\{\sum_{i=1}^{I}p_{(i)}S_{ik}Y_{ik}+\sum_{i=0}^{I}\sum_{j=1}^{J}\gamma_{ij}\nu_{ijk}\right\} \tag{43}$$

$$s.t.\ \sum_{j=1}^{L}\nu_{ijk}\le S_{ik}Y_{ik}\quad i=0\dots I\quad k=1\dots K$$

$$\sum_{i=0}^{I}\nu_{ijk}\ge D_{jk}\quad j=1\dots J\quad k=1\dots K$$

$$\nu_{ijk}\ge 0,\ Y_{ik}\in\{0,1\}$$

The above formulation follows the sample average approximation method; see e.g., Kleywegt et al. (2001). This "perfect information" version of the problem thus reduces to the repeated solution of the *deterministic* version of (M). This follows from the simple observation that (43) is completely decomposable by scenario. The deterministic version of (M) is often referred to as a capacitated network design problem, and may also be viewed as a minimum concave cost network design problem, the NP-completeness of which was shown by Lozovanu (1983). Nevertheless, solution of the deterministic problem, and hence of (PIM) is considerably simpler than that of the stochastic program (M). Large size instances are, routinely, solved with general purpose software for Mixed Integer LPs $(MILP)$, for example CPLEX.

The optimal value of (PIM) provides, of course, a lower bound for that of (M), since the latter's space of feasible solutions is a subset of that of (PIM). (A feasible solution of (PIM) is feasible for (M) if and only if it satisfies the additional constraints $Y_{ik} = Y_i$ for all $i = 1, \ldots, I$ and $k = 1, \ldots, K$, while the common selection variables $\{Y_i\}$ satisfy coverage constraint (2)). The solution of (PIM) provides an interesting benchmark for (M). It also reveals how much value may be attributed to having good forecasts of the various random input factors.

Appendix J: Alternative Constructive Improvement Heuristics

In this appendix, we outline two variants of the Greedy-Add heuristic described in Section 6. Both perform comparably to the *Greedy-Add heuristic, version 1*, but, in our numerical experiments, were slightly dominated by the latter.

Greedy-Drop Heuristic: Starting with the full set of suppliers, we eliminate, in each iteration, a supplier who can feasibly be eliminated and whose elimination results in the largest expected cost saving. The algorithm terminates at the first iteration where no such supplier exists. If Y^o denotes the set of suppliers at the start of a given iteration, $\sum_{i=1}^{I} Y_i^o \leq I$ possible eliminations of a single supplier need to be evaluated, by verifying whether the coverage constraint remains satisfied and by evaluating the function value $g(Y^o - e_i)$ for all i with $Y_i^o = 1$. To evaluate this set function $g(Y)$ for a given vector Y, we employ the basic sample averaging method: we generate a sample of K realizations from the joint distribution of the random quantities $\{S_i, D_j\}$; for each such realization, we solve the corresponding transportation problem in (3) and average the optimal objective function values. For a different approach, merely requiring the means and covariance matrix of the random variables $\{S_i, D_j\}$, we refer the reader to Natarajan et al. (2011).

A single evaluation of the set function $g(\cdot)$ reduces to the solution of a sequence of transportation problems, one for each of the realizations for the random variables $\{S_i\}$ and $\{D_j\}$. In the worst case, the algorithm's complexity consists of $\frac{1}{2}I(I+1)$ evaluations of the set function $g(\cdot)$ and the same number of tests whether the coverage constraint is satisfied.

Greedy-Add Heuristic, version 2: Identical to the previous heuristic, except for the selection of the initial set. In this version, a sample of K scenarios is generated for the joint distribution of all $\{S_i\}$ and $\{D_j\}$ variables. The full (PIM) problem is solved and the suppliers selected in at least α% of the K scenarios become part of the initial set S^o. In other words, farmer $i \in S^o$ if and only if $\frac{1}{K}\sum_{k=1}^{K} Y_{ik}^* \geq \alpha$, with Y^* an optimal solution to (PIM).

Appendix K: Model Generalizations

In this appendix, we show how our model can be generalized to incorporate various complications, frequently encountered in applications.

K.1. Salvage Value for the Manufacturer

In our base model, we have ignored the possibility of a salvage value p^s for any unit of excess supply from the farmers' contracts, above and beyond the manufacturer's aggregate demand. Such a salvage value is easily incorporated, replacing the objective function (1) in (M) by:

$$\min_{Y} \ \tilde{g}(Y) \equiv \sum_{i=1}^{I} v_i Y_i + \mathbb{E}_{\{S_i, D_j, \gamma_{0j}\}} \Psi(Y) - p^s \mathbb{E}[\sum_{i=1}^{I} S_i Y_i - \sum_{j=1}^{J} D_j]^+ \tag{44}$$

Within the $MILP$ approach of Section 6.2, the additional term can be accommodated as follows:

Choose a three-point grid $\{d_0^{'} < 0,\ d_1^{'} = 0,\ d_2^{'} > 0\}$ with $d_0^{'}$ $(d_2^{'})$ sufficiently small (large). Add new continuous variables $\{\zeta_{0k},\ \zeta_{1k},\ \zeta_{2k},\ O_k,\ k = 1, \dots, K\}$ to the $MILP$ with additional constraints:

$$O_k = \sum_{i=1}^{I} S_{ik} Y_i - \sum_{j=1}^{J} D_{jk} \qquad k = 1, \dots, K$$

$$O_k = d_0^{'} \zeta_{0k} + d_2^{'} \zeta_{2k} \qquad k = 1, \dots, K$$

$$\zeta_{0k},\ \zeta_{1k},\ \zeta_{2k} \geq 0, \quad \zeta_{0k} + \zeta_{2k} \leq 1, \quad \{\zeta_{0k},\ \zeta_{1k},\ \zeta_{2k}\} \in SOS2 \qquad k = 1, \dots, K$$

and add the term $-p^s \sum_{k=1}^{K} \delta_k d_2^{'} \zeta_{2k}$ to the objective function of (30).

The *Greedy-Add* or *Greedy-Drop* heuristics may, in principle, be applied, without any modification. Note however that, in contrast to $g(Y)$, the augmented set function $\tilde{g}(Y)$ in (44) is no longer supermodular. To maintain supermodularity, the additional term $-p^s \mathbb{E}[\sum_{i=1}^{I} S_i Y_i - \sum_{j=1}^{J} D_j]^+$ may be replaced by the modular function $-p^s \mathbb{E}[\sum_{i=1}^{I} S_i Y_i - \sum_{j=1}^{J} D_j]$. Given the coverage constraint (2), this approximation is rather innocuous.

K.2. Fixed Costs when Transacting with Farmers

In our base model, the manufacturer aims to minimize the expected value of procurement and distribution costs, subject to the coverage constraint. In terms of the procurement costs, the base model assumes these consist of the payments to the contracted farmers which are proportional to the supply quantities at unit prices specified in the selected contracts. As discussed in Section 3, there may be additional *fixed* transaction costs in contracting with any given farmer, arising, for example, from yield monitoring and quality control processes; see the Introduction. Thus, let κ_i be the fixed transaction costs when contracting with farmer i. These fixed transaction costs are effortlessly incorporated into the model, by replacing the coefficients $\{v_i\}$ in the linear term in (1) by $\hat{v}_i \equiv v_i + \kappa_i$ $(i = 1, \dots, I)$, without impacting the solution methods or any of the structural results. Clearly, as the fixed transaction costs increase, the manufacturer is likely to contract with fewer farmers, with larger supply potential, possibly less favorably located from the perspective of associated distribution costs.

K.3. Correlations Among the Supply Quantities and the Spot Price

In our base model, we have assumed that the rainfall quantities $\{R_i\}$ are independent, so that the supply variables $\{S_i\}$ are independent as well. In reality, the rainfall quantities, in particular for farmers that are geographically close to each other, are *positively* correlated. Similarly, we have assumed that the supply quantities $\{S_i\}$ are independent with respect to the future spot price which is incorporated into the cost parameters $\{\gamma_{0j},\ j=1,\dots,J\}$. However, since the aggregate rainfall impacts the spot price, we may specify

$$\gamma_{0j} = \Gamma_j(R^{tot}), \qquad j=1,\dots,J \tag{45}$$

for a given decreasing function Γ_j that can be inferred from historical data. See, for example, Table 2 in Bhanumurthy et al. (2013).

There are several ways to model a correlation structure among the rainfall random variables $\{R_i\}$. One such model would partition the country into geographical regions (e.g. districts or states in India, as in our potato procurement application). All rainfall quantities $\{R_i\}$ pertaining to farmers in the *same* region would be assumed to be dependent with a given multivariate distribution; rainfall quantities pertaining to farmers in *different* regions would be assumed to be independent.

Here is a second plausible model, with the advantage of being more parsimonious, i.e., requiring fewer model parameters. Assume the country is partitioned into R regions, with index $r=1,\dots,R$ and let $\{I_1,\dots,I_R\}$ denote the corresponding partition of the set of potential farmers $\{1,\dots,I\}$. Now, specify

$$R_i = \rho_i + \eta_r R_r^{tot}, \qquad \text{for all } i\in I_r,\ r=1,\dots,R \tag{46}$$

with $\{\rho_i\}$ a set of independent random variables, $\eta_r>0$, $r=1,\dots,R$ given constants, and $R_r^{tot}=\sum_{i\in I_r} R_i$ the aggregate rainfall among all farmers in region r.

Correlations among the yield variables $\{X_i\}$ and, hence, the supply quantities $\{S_i\}$ may arise because of other interdependencies beyond those prevailing among the rainfall quantities. For example, the manufacturer's belief distributions for the farmers' $\{\alpha_i\}$ parameters in the yield equation (9) may be positively correlated because of common climate or infrastructure factors.

Abandoning the independence assumption impacts the solution method only in *two* places. First, when evaluating the expectation in the second term of the objective functions (1) or (5) in (M) or (M'), respectively, rather than drawing a sample from independent random variables $\{S_i, D_i, \gamma_{0j}\}$, one would draw these from a joint distribution. In addition, one needs to adapt the coverage constraint. This requires calculating or estimating the covariances $\{cov(S_i,S_{i'}):\ i\neq i'\}$ for every pair of potential farmers. (Estimation is, again, most easily performed on the basis of simulation.) As demonstrated in Federgruen and Yang (2008), the following CLT-based approximation of the coverage constraint (2) can now be used:

$$\mathbb{P}\left[Y_E + U\sqrt{\sum_{i=1}^{I} Var(S_i)Y_i + \sum_{i=1}^{I}\sum_{i'\neq i} cov(S_i,S_{i'})Y_iY_{i'}} < D^{tot}\right] \le \epsilon$$

where U denotes a random variable with a standard Normal distribution. It can again be shown that this chance constraint is equivalent to the following pair of constraints, replacing (6) and (7):

$$(Y_E-\mu_{tot})^2 - z_\epsilon^2\sigma_{tot}^2 - z_\epsilon^2\left\{\sum_{i=1}^{I} Var(S_i)Y_i + \sum_{i=1}^{I}\sum_{i'\neq i} cov(S_i,S_{i'})Y_iY_{i'}\right\} \ge 0 \tag{47}$$

$$Y_E \ge \mu_{tot} + z_\epsilon\sigma_{tot}$$

The *Greedy-Add* or *Greedy-Drop* heuristics can be applied, merely using the coverage constraint (47) instead of (6). As to the $MILP$ approach in Section 6.2, various approaches may be used to transform the quadratic mixed integer program into a $MILP$. Glover (1975) showed how the quadratic terms $\{Y_i Y_{i'} : i' \neq i\}$ may be replaced by the introduction of $O(I)$ new binary variables. See Sherali and Smith (2007) for a more efficient linearization approach, and see Burer and Letchford (2012) for a survey on the topic.

As mentioned in Section 7, we have investigated how correlations among the $\{S_i\}$ variables impact for the optimal choices for the manufacturer. More specifically, we have re-evaluated the eight sets of problem instances in Table 2, with $I = 250$ and $I = 500$ farmers, in Table 13, now assuming the rainfall quantities and the $\{\alpha_i\}$ intercept parameters for farmers in a given state are drawn from a multivariate Normal distribution with the same mean and variances as in the base model, but a common correlation coefficient $\kappa = 0.25$, 0.5, and 0.75. See Equation 37 in Appendix D.2 for a full specification of the yield equation.

All problem instances are re-solved with the Greedy-Add heuristic. We have also specified the $\{\gamma_{0j}\}$ coefficients, which involve the spot price, as dependent on the *total* rainfall quantity; see Equation (40) in Appendix D.4. (The running times of the Greedy-Add heuristic are hardly affected by the increased complexity of the coverage constraint.)

In Table 13, we exhibit, for each of the eight sets of problem instances, the average value of (i) $R_2 =$ the expected aggregate supply procured, relative to the aggregate expected demand, (ii) the number of contracted farmers, and (iii) the number of farmers eliminated from or added to the optimal pool of farmers in the absence of supply correlations.

TPFS = 250				TPFS = 500			
CORR	R_2	CNF	NFEA	CORR	R_2	CNF	NFEA
0	1.223	217.0	N/A	0	1.221	445.0	N/A
0.25	1.241	218.2	17.2	0.25	1.239	449.0	31.6
0.5	1.260	218.4	18.2	0.5	1.259	458.0	32.6
0.75	1.287	225.2	20.6	0.75	1.300	470.8	45
0	1.157	198.4	N/A	0	1.155	392.8	N/A
0.25	1.169	203.4	29.4	0.25	1.166	392.4	57.2
0.5	1.182	205.8	30.2	0.5	1.175	398.8	54
0.75	1.198	211.0	30.6	0.75	1.194	409.8	59.4
0	1.221	169.4	N/A	0	1.218	320.8	N/A
0.25	1.233	167.4	26.4	0.25	1.232	319.6	59.2
0.5	1.247	170.8	28.2	0.5	1.243	330.4	66
0.75	1.263	174.4	26.6	0.75	1.259	335.2	73.6
0	1.156	161.2	N/A	0	1.155	270.6	N/A
0.25	1.165	159.2	29.2	0.25	1.161	267.6	37.4
0.5	1.173	163.0	30.2	0.5	1.170	267.6	39
0.75	1.186	165.2	29.2	0.75	1.177	274.2	47.2

Table 13 (Near-)Optimal manufacturer decisions for the correlated-supply scenarios

TPFS: Targeted Potential Farmer Size, CORR: Rainfall/α_i correlation coefficient, CNF: Contracted Number of Farmers, NFEA: Number of Farmer Eliminated/Added compared to the independent-supply (CORR=0) scenario.

Clearly, as the correlation among the supply quantities increases, it is optimal for the manufacturer to increase her expected aggregate contracted supply: as the correlation coefficient increases, so does the variability of the aggregate supply under any given set of contracted farmers, and associated contracts, requiring an increase of R_2 to continue to satisfy the coverage constraint. Indeed, we observe a consistent increase of the R_2-value as we increase the correlation coefficient: as an example, when there are 500 potential farmers, a high demand scenario is faced, and a service level of 99% is targeted, it is optimal to increase the expected "safety supply" from 22% to 30% of the expected aggregate demand.

In general, the increased aggregate volatility causes the manufacturer to contract with more farmers, but this is not always the case. The increased correlation patterns result in significant changes in the pool of the contracted farmers, resulting from farmer eliminations as well as additions. Consider, again, the above set of problem instances with 500 potential farmers: on average, 45 modifications are made to the set of (on average) 445 contracted farmers that is optimal in the absence of supply correlations. In other words, the number of farmer additions and eliminations is approximately 10% of the size of the optimal farmer pool in the absence of supply correlations. Indeed, the optimal set of farmers in the base case with $\kappa = 0$ often fails to be feasible when $\kappa > 0$, because the last term in the coverage constraint (47) becomes more negative. The manufacturer therefore needs to substitute some farmers, contracting with more reliable but perhaps less conveniently located farmers.

We have repeated the numerical study of Table 3 assuming the supply variables are correlated, adopting the same yield model as in Table 13, again with $\kappa =$0.25, 0.5, and 0.75. The results are reported in Table 14. Recall that the objective of this study is to assess the impact of the contract menu on various performance measures. As can be expected, we observe that the manufacturer's cost uniformly increases with the correlation coefficient κ, when comparing otherwise identical instances and contract menus. The increased cost is the consequence of the volatility of aggregate supplies increasing with κ, hence requiring a larger expected supply volume to comply with the coverage constraint. Focusing on the generally most effective Slope 3 and 3-contract options menu, aggregate costs increase by 7-8% when changing from $\kappa = 0$ (independent supplies) to $\kappa = 0.25$.

A similar pattern can be observed when increasing κ from $\kappa = 0.25$ to $\kappa = 0.5$ or from $\kappa = 0.5$ to $\kappa = 0.75$. The increased volatility of aggregate supplies due to positive correlations among the supply variables, also necessitates increased diversification among a larger pool of farmers. Focusing, again, on the same Slope 3 and 3-contract options menu and instances with triple the rainfall standard deviations compared to the base case, we observe that in the absence of correlations, it is optimal to contract with 129 farmers; when $\kappa = 0.25$, a net addition of 10 additional farmers is required. Yet another net addition of 11 farmers is called for, to a total of 150 farmers, when κ is increased from $\kappa = 0.25$ to $\kappa = 0.5$. If κ is further increased to $\kappa = 0.75$, there is no feasible set of farmers among the 158 potential farmers. More generally, the set of problem instances for which no feasible solution is found increases with κ. When $\kappa = 0$, only the instances with the highest demand and rainfall volatilities (demand CV = 0.3 and rainfall CV = 3x) are infeasible and only under Slope 1 with three options. When $\kappa = 0.25$, the same set of instances is infeasible, under either the Slope 1 or Slope 2 menus and regardless of whether the menu offers three or five options. When $\kappa = 0.5$, no feasible solution exists even under Slope 3; moreover, even the instances in the second segment, i.e., with demand CV = 0.15 and rainfall CV = 3x, become infeasible as long as Slope 1 is used. Finally, when $\kappa = 0.75$ and the standard deviation of the rainfall quantities are triple their values in the base case, more than $I = 158$ farmers are needed to comply with the coverage constraint, regardless of which of the menus is chosen.

		MC		CNF		Percentage of CNF by Contract Option (%)							
# of Contract Options:		Five	Three	Five	Three	Five					Three		
						O1	O2	O3	O4	O5	O1	O3	O5
D. CV = 0.15	Slope 1	615,422	585,104	125	124	13.6	33.6	19.2	31.2	2.4	24.2	60.5	15.3
R. CV= x	Slope 2	570,479	571,343	118	119	11.9	10.2	39.8	33.9	4.2	11.8	58.8	29.4
CORR = 0.25	Slope 3	555,258	580,813	117	117	5.1	12.8	29.9	36.8	15.4	7.7	53.8	38.5
D. CV = 0.15	Slope 1	701,960	723,665	148	154	10.1	30.4	33.1	23.6	2.7	31.2	49.4	19.5
R. CV = 3x	Slope 2	636,349	656,285	140	142	5.7	8.6	44.3	37.9	3.6	7.0	66.9	26.1
CORR = 0.25	Slope 3	628,558	634,220	138	139	5.8	5.8	31.9	39.1	17.4	5.8	46.0	48.2
D. CV = 0.3	Slope 1	716,469	734,229	141	141	12.1	30.5	19.9	34.0	3.5	21.3	63.1	15.6
R. CV = x	Slope 2	700,932	687,520	135	135	10.4	8.9	35.6	41.5	3.7	10.4	53.3	36.3
CORR = 0.25	Slope 3	679,258	686,771	133	134	4.5	12.0	30.1	35.3	18.0	6.7	51.5	41.8
D. CV = 0.3	Slope 1	Infeasible											
R.CV = 3x	Slope 2												
CORR = 0.25	Slope 3	790,454	776,091	156	155	5.1	5.1	30.1	38.5	21.2	5.2	41.9	52.9
D. CV = 0.15	Slope 1	621,309	618,160	126	126	13.5	33.3	18.3	31.7	3.2	23.8	61.1	15.1
R. CV= x	Slope 2	579,238	571,403	120	121	11.7	10.8	37.5	37.5	2.5	11.6	57.9	30.6
CORR = 0.5	Slope 3	574,743	593,727	119	119	5.0	12.6	29.4	38.7	14.3	7.6	55.5	37.0
D. CV = 0.15	Slope 1	Infeasible											
R. CV = 3x	Slope 2	724,454	743,681	151	154	5.3	7.9	42.4	41.1	3.3	6.5	66.2	27.3
CORR = 0.5	Slope 3	693,241	719,425	147	150	5.4	5.4	31.3	38.8	19.0	5.3	42.0	52.7
D. CV = 0.3	Slope 1	723,082	706,161	143	143	11.9	30.8	17.5	37.1	2.8	21.0	59.4	19.6
R. CV = x	Slope 2	678,131	687,023	136	137	10.3	8.8	35.3	41.9	3.7	10.2	55.5	34.3
CORR = 0.5	Slope 3	678,380	675,760	135	135	4.4	11.1	28.1	36.3	20.0	6.7	51.1	42.2
D. CV = 0.3	Slope 1	Infeasible											
R.CV = 3x	Slope 2												
CORR = 0.5	Slope 3												
D. CV = 0.15	Slope 1	659,699	623,538	129	128	13.2	33.3	20.2	30.2	3.1	23.4	64.1	12.5
R. CV= x	Slope 2	593,387	604,791	122	122	11.5	9.0	39.3	37.7	2.5	11.5	56.6	32.0
CORR = 0.75	Slope 3	583,136	581,053	120	121	5.0	12.5	29.2	37.5	15.8	7.4	52.1	40.5
D. CV = 0.15	Slope 1	Infeasible											
R. CV = 3x	Slope 2												
CORR = 0.75	Slope 3												
D. CV = 0.3	Slope 1	735,258	729,021	145	144	11.7	30.3	19.3	35.9	2.8	20.8	61.8	17.4
R. CV = x	Slope 2	714,031	704,429	138	139	10.1	8.0	36.2	42.0	3.6	10.1	54.0	36.0
CORR = 0.75	Slope 3	685,307	714,659	136	136	4.4	10.3	28.7	36.0	20.6	6.6	50.0	43.4
D. CV = 0.3	Slope 1	Infeasible											
R.CV = 3x	Slope 2												
CORR = 0.75	Slope 3												

Table 14 Impact of menu design on manufacturer's performance measures for the correlated-supply scenarios D.CV: Demand Coefficient of Variation, R.CV: Rainfall Coefficient of Variation, CORR: Rainfall/α_i correlation coefficient, MC: Manufacturer Cost, CNF: Contracted Number of Farmers, Oi: Contract Option i

References

Bhanumurthy, N. R., P. Dua, L. Kumawat. 2013. Weather shocks and agricultural commodity prices in India. *Climate Change Economics.* **4**(3).

Burer, S., A. N. Letchford. 2012. Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science* **17** 97–106.

Glover, F. 1975. Improved linear integer programming formulations of nonlinear integer problems. *Management Sci.* **22** 455–460.

Hull, J. 2013. Options, Futures, and Other Derivatives (9th Edition). Pearson Inc., NJ.

Kleywegt, A. J., A. Shapiro, T. Homem de Mello. 2001. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* **12** 479–502.

Lozovanu, D. 1983. Properties of optimal solutions of a grid transport problem with concave function of the flows on the arcs. *Engineering Cybernetics* **20** 34–38.

Natarajan, K., C. P. Teo, Z. Zheng. 2011. Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Oper. Res.* **59**(3) 713–728.

Sherali, H. D., J. C. Smith. 2007. An improved linearization strategy for zero-one quadratic programming problems. *Ptim. Lett.* **1**(1) 35–47.

Singh, R. B., P. Kumar, T. Woodhead. 2002. Smallholder farmers in india: Food security and agricultural policy. FAO Report, FAO Regional Office for Asia and the Pacific, Bangkok, Thailand.

Topkis, D. M. 1998. Supermodularity and Complementarity. Princeton University Press, Princeton, NJ.