For Online Publication

Appendix A: Hiding Output

In this appendix we consider a variant of the basic model, in which effort is costless, but the agent may hide output. In particular, the agent may report that output is low even when it is high. The principal provides the agent with a bonus a if reported output is high, but may dismiss the agent ($d = 1$) if the reported output is low and the signal indicates that the state of nature is good. The basic wage in this case covers subsistence: $\omega = m$.

An incentive scheme, $a > 0, d \in \{0, 1\}$, induces truthful reporting of the agent if:[1]

$$a + \delta V \geq (H - L) + ((q(1 - d) + (1 - q))\delta V. \tag{A1}$$

where $H - L$ is the output stolen by the agent when he reports low instead of high output, and V denotes the present value of the agent's utility from being employed in agriculture in a stationary equilibrium with truthful reporting. The agent's incentive constraint is binding in the optimal solution (otherwise the principal can lower the bonus payment a) and so:

$$a = (H - L) - q\delta dV. \tag{A2}$$

The value function $V(a, d)$ associated with truthful reporting (analog of (2) in the basic model) is:

$$V(a, d) = \frac{pa}{1 - \delta(1 - \mu d)}. \tag{A3}$$

Plugging (A3) into (A2) and simplifying yields an incentive constraint:

$$a = (H - L)\left(1 - \frac{\delta pqd}{1 - \delta + \delta d\,(\mu + pq)}\right). \tag{A4}$$

[1]Notice that the incentive constraint is relevant only in case the state of nature is good and output is high.

The principal's objective is:

$$\pi = \max_{a,d\in\{0,1\}} p(H-L) + L - pa - \mu dx - m, \tag{A5}$$

subject to (A4).

Thus, two types of contracts may be optimal: one with $d = 0$ ('pure carrot') and another with $d = 1$ ('carrot and stick'). The threshold transparency level $\hat{q}$ that determines the level above which the'carrot and stick' is optimal is given by the solution of the following equation (analogous to (4) in the basic model) that equates the expected profit to the principal under the two contracts:

$$\frac{\hat{q}}{1-\hat{q}} = \frac{(1-p)\,x}{p\delta(H-L)}[1-\delta(p+\hat{q}\ -2p\hat{q})]. \tag{A6}$$

A pure carrot contract is optimal if $q < \hat{q}$. It is given by:

$$d_c = 0, a_c = H-L, \text{ and } V_c = p(H-L)/(1-\delta). \tag{A7}$$

A stick and carrot contract is optimal if $q > \hat{q}$. It is given by:

$$d_s = 1, a_s = (H{-}L)\left(1-\frac{\delta pq}{1-\delta(p+q-2qp)}\right), \qquad V_s = \frac{p(H-L)}{1-\delta(p+q-2qp)}. \tag{A8}$$

These results reveal that the analysis of the main model is qualitatively robust to this alternative scenario of the moral hazard problem.

Appendix B: Costly Monitoring

Suppose that the model is identical to the basic model except that the principal can observe a signal $\sigma \in \left\{\tilde{l}, \tilde{h}\right\}$ about the agent's effort at cost $c \geq 0$ (in units of output) instead of on the state of nature as in the basic model. The accuracy of the signal is $q \in [1/2, 1]$, such that:

$$Pr(\tilde{h}|h) = Pr(\tilde{l}|l) = q \ \ ; \ Pr(\tilde{h}|l) = Pr(\tilde{l}|h) = 1-q.$$

The case of a perfect monitoring is captured by: $q = 1$; and the case where it is uninformative is captured by: $q = 1/2$.

As in the basic model, $\gamma > 0$ is the periodic cost of exerting high effort, the agent's alternative employment outside of agriculture tenancy provides utility of zero and the agent's periodic utility, U, when engaged in agriculture equals his expected income, to be denoted by I, less the cost of effort. In particular, when exerting high effort, this periodic utility is: $U = I - \gamma$.

We denote the present value of the agent's utility from being employed in agriculture by V, and denote by $\delta \in (0, 1)$ the agent's discount factor.

The principal is assumed to rely on the following incentive scheme. If output is high, then the principal retains the agent with certainty and pays the agent $\omega + a$, where $a \geq 0$ is a bonus payment. If output is low, then the agent is still paid the basic subsistence wage $\omega = \gamma$.

When output is low, if the signal indicates that the agent was exerting high effort ($\sigma = \tilde{h}$), then the principal retains the agent. But if output is low and the signal indicates that the agent was shirking ($\sigma = \tilde{l}$), then the principal may dismiss the agent.

We denote by $d = 1$ the strategy of dismissal upon low output and a signal indicating low effort: $\sigma = \tilde{l}$ and $Y = L$, and retention of the agent otherwise, and by $d = 0$ the strategy of always retaining the agent. If the agent is dismissed, the principal incurs a fixed cost $x > 0$ (in units of output). We assume that this cost is large enough to ensure that it will not be desirable to dismiss the agent when output is low ($Y = L$) and the signal indicates high effort.

Thus, the principal can either imply a contract with $d = 1$ in which he incurs the monitoring cost c, or she can employ a contract with $d = 0$ and no monitoring.

Given our normalization that the utility of a dismissed agent is zero, in a stationary equilibrium the value of the employed agent's discounted utility, when he exerts high effort, has to satisfy:

$$V = pa + [1 - Pr(dismiss|e = h)]\delta V. \tag{B1}$$

For convenience, we denote the probability of a bad harvest and a good signal by $\mu = (1-p)(1-q)$. The probability of dismissal upon high effort is then $d\mu$. V is thus determined by the contract parameters a and d and the parameters: μ, p and δ as follows:

$$V(a,d) = \frac{pa}{1 - \delta(1 - \mu d)}. \tag{B2}$$

The principal's objective is to solve for the employment contract that maximizes her periodic expected payoff, denoted by π,

$$\pi = \underset{a \geq 0, d \in \{0,1\}}{max} \; p(H-a) + (1-p)L - \mu dx - \omega - dc,$$

subject to providing the agent with incentives to exert high effort (identical to the basic model):

$$\begin{aligned} & p(a + \delta V) + (1-p)[q + (1-q)(1-d)]\delta V + \omega - \gamma \\ & \geq \\ & p(q(1-d) + (1-q))\delta V + (1-p)[(q + (1-q)(1-d)]\delta V + \omega, \end{aligned}$$

where $V = V(a,d)$ as in (B2).

Since $\omega = \gamma$, we can rewrite the principal's objective function and the agent's incentive constraint as follows:

$$\pi = \max_{a \geq 0, d \in [0,1]} p(H - L) + L - \gamma - pa - \mu dx - dc, \tag{\textit{B-OF}}$$

s.t.

$$pa + pqd\delta V(a,d) \geq \gamma. \tag{\textit{B-IC}}$$

Thus, we obtain that modeling monitoring as a (costly) signal on effort, yields a maximization problem that for $c = 0$ is identical to the maximization problem in the main model. More generally, the larger is c the higher would be the threshold $\hat{q}$ above which the optimal contract is 'stick & carrot', without any change in the qualitative results. This indicates that the larger is c - the more costly it is to obtain a signal on effort as in this model or on the state of nature, as in the main model - the larger is the range of parameters for which the solution is 'pure carrot'. This means that if $c > 0$ then the threshold $\hat{q}$ is strictly larger than $1/2$ for lower values of the cost of replacement x.

Appendix C: Probabilistic Dismissal

In this appendix we consider again the basic model, but we allow the principal to dismiss the agent upon observation of low output and a good signal with any probability $d \in [0,1]$ as opposed to just $d \in \{0,1\}$ as in the main text. We recast the principal's problem as the minimization of discretionary expenditure:

$$\min_{d\in[0,1],a} pa + \mu xd, \tag{C1}$$

subject to the agent's incentive constraint:

$$pa = \left(1 + \frac{pqd\delta}{1-\delta(1-d\mu)}\right) \geq \gamma. \tag{C2}$$

The agent's incentive constraint must be binding in the optimal solution. Plugging the value of a from (C2) into (C1) yields the principal's objective function

$$\gamma\left(1 - \frac{\delta pqd}{1-\delta+\delta d\left(\mu+pq\right)}\right) + \mu xd. \tag{C3}$$

as a function of d alone.

Differentiation of the principal's objective function with respect to d yields:

$$-\frac{\gamma\delta qp(1-\delta)}{A^2} + \mu x \tag{C4}$$

where $A = 1 - \delta + \delta d(\mu + pq)$.

Inspection of (C3) reveals that the expression on the left of (C3) is convex in d while the expression on the right is linear and increasing in d. Comparison of the values of these two expressions at $d = 0$ reveals that if

$$q \leq \frac{(1-p)x(1-\delta)}{\delta\gamma + (1-p)x(1-\delta)} \tag{C5}$$

then the value of d that maximizes the principal's objective function (sets the derivative (C4) equal to zero) is negative. Because d is a probability, this means that the optimal probability of dismissal in this case is $d = 0$. Comparison of the values of these two expressions at $d = 1$ yields another condition on q such that the value of d that maximizes the principal's objective function is larger than one. Because d is a probability, this means that the optimal probability of dismissal in this case is $d = 1$.

Thus, there exist two threshold values $\underline{q}$ and $\overline{q}$ such that for $q < \underline{q}$ the optimal $d = 0$; for $q > \overline{q}$ the optimal $d = 1$; and for $\underline{q} \leq q \leq q < \overline{q}$ the optimal value of d (obtained from solving the first-order-condition equation $C4 = 0$) is given by:

$$d = \frac{1-\delta}{\delta(\mu + pq)} \left(\sqrt{\frac{\gamma\delta pq}{(1-\delta)\mu x}} - 1 \right) \tag{C6}$$

If the right-hand-side of (C5) is larger than .5 or, equivalently,

$$\frac{(1-p)x}{\gamma} > \frac{\delta}{1-\delta} \tag{C7}$$

then $\underline{q} > .5$, which means that the pure carrot contract is optimal for some values of the accuracy parameter q. Inspection of (C7) reveals that this is the case if the cost of dismissal x is sufficiently large and/or the agent is impatient (δ is small) so that the threat of dismissal is less effective.

The next figure depicts the optimal dismissal probability d as a function of transparency q for the same parameters as in the example in the main text:

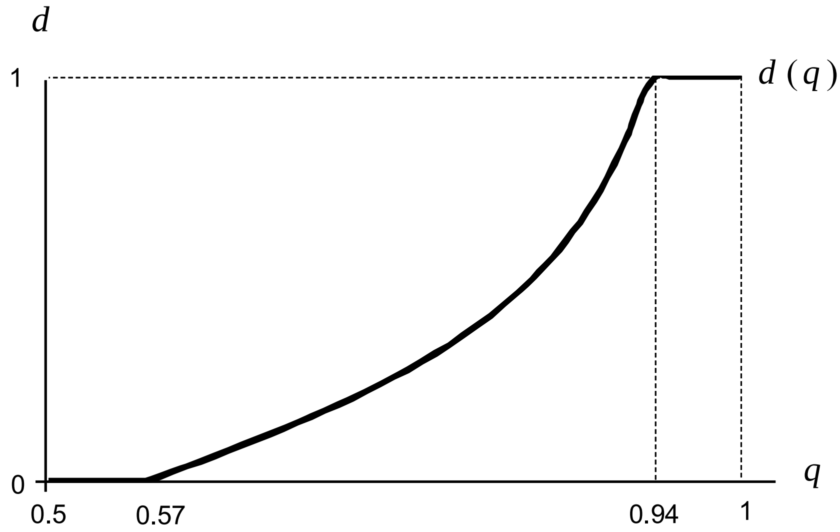

Figure 5: The optimal dismissal probability, $d \in [0,1]$, as a function of transparency q

As in the basic case, the agent's bonus is maximal when $q < \underline{q}$. In the range above $\underline{q}$, as the probability of dismissal increases, the bonus decreases – since the increased threat of dismissal is used as a substitute incentive device. The bonus continues to decrease further in the range where $q > \overline{q}$, where the dismissal probability reaches its upper limit ($d = 1$). The principal's net expected revenue (taking into account the costs of dismissal) is constant below the threshold $\underline{q}$ and increases monotonically in q above $\underline{q}$.

Appendix D: Warning before Dismissal

In this appendix we allow the principal to warn the agent an optimally chosen number of times when output is low and the signal about the state of nature is good before actually dismissing the agent. That is, we assume that the principal optimally selects an integer number n of "bad signals," or times at which will observe $Y = L$ and $\sigma = g$ before it dismisses the agent.

The number of "warnings" prior to dismissal is thus given by $n - 1$. The basic model is therefore one where n is restricted to the set $\{1, \infty\}$.

Let $V(n)$ denote the value of being employed in agriculture for an agent with n bad signals left. If $n = 1$ then the agent is dismissed the next time $Y = L$ and $\sigma = g$. The agent is dismissed immediately upon $n = 0$ and so $V(0) = 0$. Let $a(n)$ denote the bonus payment to the agent when $Y = H$ as a function of the number of bad signals that remain n.

The value function $V(n)$ satisfies the following recursive equation:

$$V(n) = pa(n) + \mu\delta V(n-1) + (1-\mu)\delta V(n). \tag{D1}$$

The agent's incentive constraint, which as before is binding in the optimal solution, can be simplified to:

$$pa(n) = \gamma - pq\delta(V(n) - V(n-1)). \tag{D2}$$

By combining (D1) and (D2) we obtain the following recursive formulation for $V(n)$:

$$V(n) = A + BV(n-1), \tag{D3}$$

where the constants A and B are given by:

$$A = \frac{\gamma}{1 - \delta + \delta(\mu + pq)}; B = \frac{\delta(\mu + pq)}{1 - \delta + \delta(\mu + pq)}. \tag{D4}$$

Observe that $0 < A$ and $0 < B < 1$.

Given that $V(0) = 0$, the solution for $V(n)$ in terms of the parameters of the model is:

$$V(n) = \frac{A(1 - B^n)}{1 - B}. \tag{D5}$$

It therefore follows that:

$$a(n) = \gamma/p - q\delta AB^{n-1}. \tag{D6}$$

Observe that the bonus payments to the agent increase with n. It can be immediately verified that $a(1)$ and $V(1)$ are identical to a_s and V_s of the basic model, while a_c and V_c coincide to the limits of $a(n)$ and $V(n)$ from (D6) and (D5), respectively, as n tends to infinity.

We now solve for the optimal number n. Denote the principal's discount factor by δ_P, and denote the discounted expected discretionary costs for the principal (that include bonus payments and dismissal costs) starting from the point where it employs an agent has k bad signals left until dismissal under a policy where agents are dismissed after n bad signals and are induced to exert high effort in every period by $c(k, n)$.

For $k = 1$:

$$\varphi(1, n) = pa(1) + \mu(x + \delta_P c(n, n)) + (1 - \mu)\delta_P c(1, n).$$

And for $1 < k \leq n$:

$$c(k, n) = pa(k) + \mu\delta_P c(k - 1, n)(1 - \mu)\delta_P c(k, n).$$

These two equations simplify to:

$$c(1, n) = \alpha a(1) + \beta x/\delta_P + \beta c(n, n), \tag{D7}$$

and

$$c(k, n) = \alpha a(k) + \beta c(k - 1, n), \tag{D8}$$

where the two constants α and β are given by:

$$\alpha = \frac{p}{1 - \delta_P + \mu\delta_P}; \quad \beta = \frac{\mu\delta_P}{1 - \delta 2 + \mu\delta_P}. \tag{D9}$$

Equations (D7) and (D8) can be explicitly solved for $c(n, n)$ as a function of the underlying parameters of the model as follows:

$$c(n, n) = \frac{\gamma}{1 - \delta_P} + \frac{\beta^n x}{\delta_P(1 - \beta^n)} + \frac{\alpha q \delta A(B^n - \beta^n)}{(1 - \beta^n)(B - \beta)}. \tag{D10}$$

It is reassuring to confirm that the solution of the equation $c(1,1) = c(\infty, \infty)$ for q yields the threshold $\hat{q}$ from the basic model, and is independent of the principal's discount factor δ_P.

The following figure describes the optimal n (the n that minimizes (D10)) as a function of the level of transparency q, for the same parameters used to illustrate the basic model. The additional parameter δ_P is set to $\delta_P = 0.98$.[2]

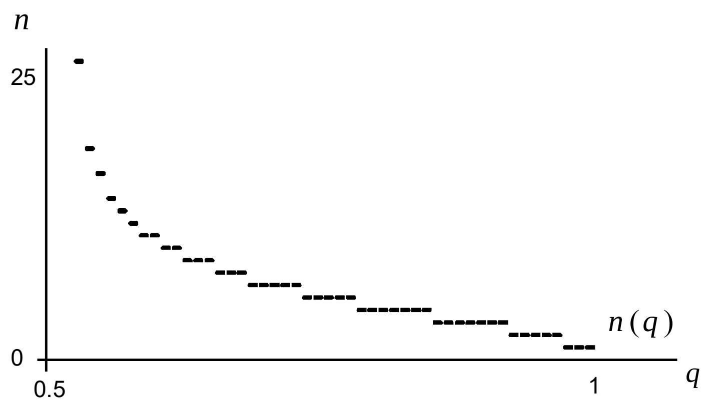

Figure 6: The optimal number of "bad signals" before dismissal, n, as a function of transparency q

This analysis confirms the robustness of our basic results. There may be a range with sufficiently low transparency where permanent tenancy is provided. In this range, the total cost to the principal is highest and the bonus payments are maximal. As transparency increases, the optimal n decreases. In this range, as the information improves, the principal relies more and more on the threat of dismissal to incentivize the agent (in the sense of providing a smaller number of warnings) and at the same time also provides lower bonuses. Thus, once again opacity of production provides the tenant with both a form of de-facto property rights and greater reward for exerting effort.

[2] A lower discount rate for the principal reduces the discounted cost of dismissal and shifts the curve of optimal n's downwards.

Finally, it should be noted that in our calibration the probability of a bad signal (upon exerting effort) is $\mu = 0.2(1-q)$. Hence, a bad signal or warning is not issued more frequently than about every five years. In this case, the expected time needed for five warnings is much larger than the expected life span of an adult farmer, and so is effectively equal to infinity.

Appendix E: Endogenous Population Size

In this appendix we allow the principal to control the size of individual plots. This generalization yields new predictions with respect to the effect of transparency on the size of the population.

Suppose that output from a plot of size λ is:

$$Y(\lambda) = \begin{cases} \lambda H & \text{if } e = h \text{ and } \theta = G; \\ \lambda L & \text{otherwise.} \end{cases}$$

The agent's cost of high effort is denoted by $\gamma(\lambda)$. The cost function $\gamma(\lambda)$ is assumed to be increasing and convex and to be such that $\gamma(0) = 0$. A larger plot size is associated with a larger cost of training a new agent. We therefore assume that the replacement loss is given by $x(\lambda) = \lambda x$.

If the size of the land is controlled by the principal is T, then the number of plots (and agents) is given by T/λ. The principal is assumed to maximize her expected payoff from the entire land under her control. Thus, her problem is:

$$\Pi = \max_{\lambda>0, a\geq 0, d\in\{0,1\}} (T/\lambda)[p(\lambda H - \lambda L) + \lambda L - \omega - pa - (1-q)d\lambda x],$$

s.t.

$$pa + qd\delta V \geq \gamma(\lambda),$$

$$\omega \geq m + \gamma(\lambda).$$

The analysis of the basic model where $\lambda = 1$ applies to any $\lambda > 0$. Both the subsistence and incentive constraints are binding in the optimal solution, which implies that $\omega = m + \gamma(\lambda)$. If the signal about the state of nature is uninformative (q is sufficiently low), a 'pure carrot' contract where:

$$d_c = 0, \qquad a_c = \gamma(\lambda)/p \tag{E1}$$

is optimal. The principal's problem in this range is equivalent to the selection of λ to minimize $T(m + 2\gamma(\lambda))/\lambda$. Given the convexity of $\gamma(\lambda)$, the optimal λ_c is given by the unique solution to the first order condition:

$$\lambda_c\gamma'(\lambda_c) - \gamma'(\lambda_c) = \frac{m}{2}. \tag{E2}$$

Similarly, if the signal about the state of nature is sufficiently informative (q is sufficiently high), then a 'stick and carrot' contract where:

$$d_s = 1, \qquad a_s(q, \lambda) = \frac{\gamma(\lambda)}{p} - \frac{q\delta\gamma(\lambda)}{1 - \delta(p + q - 2pq)}. \tag{E3}$$

is optimal. The principal's problem in this range is equivalent to the selection of λ to minimize $T(m + \gamma(\lambda) + pa_s(q, \lambda)]/\lambda$. As before, the optimal solution λ_s is given by the unique solution to the first order condition:

$$\lambda_s\gamma'(\lambda_s) - \gamma'(\lambda_s) = \frac{m}{2 - \frac{pq\delta}{1-\delta(p+q-2pq)}}. \tag{E4}$$

The convexity of $\gamma(\lambda)$ implies that the left-hand-side of (E2) and (E4) is increasing in λ. The fact that the right-hand-side of (E2) is smaller than that of (E4) and the right-hand-side of (E4) is increasing in q implies that the optimal plot size under the 'stick and carrot' regime λ_s increases with transparency q, and is larger than the optimal plot size under the 'carrot' regime λ_c.

The fact that $\lambda_s > \lambda_c$ is due to the fact that when the stick is in use, it costs less to incentivize the agent, and so the principal may as well assign a larger plot size to the agent, which would allow it to economize on the fixed cost of agents' maintenance. The larger plot size implies, of course, a smaller population.

The extra decision variable λ leads to a higher expected revenue to the principal, in comparison with the case of a fixed plot size. To better evaluate the impact of endogenous plot size, consider the case where the cost function $\gamma(\lambda)$ has a constant elasticity $\lambda\gamma'(\lambda)/\gamma(\lambda) = K$, calibrated so that $\gamma(1) = \gamma$ so that the optimal plot size under the 'pure carrot' regime is still equal to one ($\lambda_c = 1$). This guarantees that under the 'pure carrot' contract every aspect of the economy is identical to that of an economy with a fixed plot size. However, the higher revenue under the 'stick and carrot' regime implies that the new threshold transparency $\hat{q}_\lambda$ for switching into the 'stick and carrot' contract is lower than before. At the transparency threshold $\hat{q}_\lambda$ the agents are made discretely worse off when they are switched from a 'pure carrot' contract to a 'stick and carrot' contract. But beyond this point, since each agent's net per-period utility depends positively on the expected bonus payment pa for high effort, the larger plot size implies that agents are made better off as transparency increases. Moreover, beyond the old threshold level $\hat{q}$ agents are better off than under the fixed plot case. This is compatible with increased revenue to the principal, since the number of agents is smaller.

These results are similar to those depicted in Figure 1. If we set $T = 1$ so that the principal's expected income is identical to her income under a fixed plot size, then the threshold $\hat{q}_\lambda$ is smaller and the principal's income above the threshold is higher. It should be noted that in a figure that captures the principal's income when plot size is endogenous the vertical difference between the two lines does not represent each agent's expected income, since this (as noted above) is in fact increasing, due to the larger plot size.

To conclude, this appendix shows that if plot size is endogenous then as economic activity becomes more transparent, the lower is population density.

Appendix F: The Urban Sector

In the model, we implicitly assume that all those individuals who do not belong to the elite and are not employed in agriculture belong to the urban sector. To simplify, we assume further that the urban sector does not trade with the farming sector. That is, the provision of protection and the collection of tribute ('protection' revenue) is the only interaction between the two sectors. We also simplify by consideration of a model with a single tier of government, where the governor is identical to the king. The food collected by the governor is evidently not consumed entirely by her. This food revenue provides the means for supporting an army that provides protection to the farming sector and secures the governor's monopoly on the extraction of revenue from farming activity. This food supply also sustains the artisans who supply various amenities (including luxury items) for the governor and his dependents, and may also possibly be exchanged for prestige goods from abroad. Since some of the food that reaches the urban sector is in some sense wasted on sumptuary meals or on imports, the ratio of the average food collection to the food required for long-term maintenance of farmers (m) provides an estimate of an upper bound on the size of the urban sector that is supported by the farming sector.[3]

More significant than the relative sizes of the two sectors is the very different uncertainty in food supply that they face. The essence of this issue can be clarified by considering what happens in bad years. At the level of the individual farmer bad years occur with probability $1 - p_1 p_2$. At the governor's level, however, they occur less frequently, with a lower probability of $1 - p_2$. This reflects the fact that the governor's revenue bundles together the revenue from many independent plots, and thus provides an insurance against idiosyncratic plot bad states. However, our model also identifies a difference in the severity of bad harvests due to village bad states. In this case, our assumptions imply that the output of each farmer is L_1, and the revenue collected by the governor is $L_2 = N_1 \left[L_1 - (m_1 + \gamma)\right]$. In the numerical calibration presented in the main text we set $L_1 = m_1 + \gamma$. This implies that the income retained after a bad harvest enables farmers to survive until

[3] If farmers are employed in the construction of monuments over the Summer, and are paid for their extra effort by the state, as was customary in Egypt, this too would have to be taken into account.

the next harvest, but the governor and the urban sector obtain no revenue at all. This extreme result is clearly due to our simple model and to this particular calibration; but it reflects a general phenomenon: a larger share of the farming output remains in the periphery after bad harvests. This captures another important and ill-understood aspect of ancient economies in which the urban sector was likely to be more vulnerable to downward shocks to output. This implies that hunger and starvation are likely to be concentrated particularly among the lower strata of the urban sector: servants, small artisans and the like. This implication is in line with our presumption that this segment of society is demographically vulnerable, and may not have reproduced on its own, other than through an inflow from the farming sector. In addition, under the circumstances assumed here, the vulnerability of the urban sector implies that whereas farmers need only store food within the year, inter-annual storage is an absolute necessity for the urban sector, as a buffer for years where the harvest is small. This inter annual storage, however, should not be considered as providing insurance for the farming sector, but rather as serving the urban sector.[4]

[4]This conclusion is consistent with the predominant archaeological finding of storage pits and granaries in ancient urban centers, but is inconsistent with the common presumption (see for example Adams (1981, p. 244; 2005)), that urban central storage served the entire population and was possibly the main service that the state provided to the countryside.