Three Essays on Institutions and Economic Development

by

Kaivan D. Munshi

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1995

Copyright Kaivan D. Munshi, 1995. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly
paper and electronic copies of this thesis document in whole or in part, and to grant
others the right to do so.

Author

Department of Economics

May 19, 1995

Certified by

Abhijit V. Banerjee
Associate Professor of Economics
Thesis Supervisor

Accepted by

Richard S. Eckaus
Chairman, Departmental Committee on Graduate Students

ARCHIVES

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 12 1995
Three Essays on Institutions and Economic Development

Kaivan D. Munshi

Submitted to the Department of Economics on May 19, 1995, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Abstract

The role of institutions in the economic development process has been the subject of considerable research attention in recent years. This dissertation comprises three empirical essays which explore specific institutions, in the development context.

The first essay, "Investment Behavior, Financial Structure and Corporate Control: The Case of the Japanese Pharmaceutical Industry," explores the relationship between financial structure and investment behavior, conditional on corporate control within the firm. The Japanese financial system is characterized by a close firm-bank relationship and a traditional reliance on bank debt as the chief source of external finance. Evidence is presented in favor of the proposition that a close firm-bank relationship reduces information problems in Japanese capital markets. Corporate control appears to respond to changes in the economic and financial environment, both over time as well as across firms, providing empirical support for the corporate governance hypothesis as the mechanism through which such information problems are reduced. Furthermore, a close firm-bank relationship does not appear to distort the investment-portfolio toward excessively safe projects in the Japanese pharmaceutical industry, which is a potential consequence of the bank's concave payoff function under the standard debt contract. From a development perspective, these are encouraging results, in light of recent attempts to replicate the Japanese financial system in developing countries and transforming socialist economies.

The second essay, "Milk Supply Behavior in India: Data Integration, Estimation and Implications for Dairy Development," studies the role of village-level milk producers' cooperative societies in increasing production in India's dairy industry. We develop and estimate a model to derive the sources of growth in Indian milk production. A substantial share of this growth is attributed to technological progress associated with the cooperative system. Directed interventions in the dairy sector can account for only a fraction of this technological progress. For instance, cross-breeding only partially explains the growth in cow milk production, suggesting improvement in the quality of indigenous cows. Diffuse determinants may thus play an important role in dairy development.
In this regard, the cooperative system provides a channel for the dissemination of information as well as an infrastructure base for the adoption of new technology.

The third essay, "Social Learning and Technology Diffusion: An Application to Indian Agriculture," attempts to place structure on the dissemination of information, described above. This essay does not deal explicitly with institutions, focussing instead on individual peasant-behavior and the diffusion of Green Revolution technology in Indian agriculture. It nevertheless presents a mechanism for the transmission of information which may help to explain institutional dynamics in developing countries. Technology diffusion is interpreted in this essay as the outcome of a social learning process in which an agent learns from his neighbors about the quality of a new and uncertain technology. Alternative models of social learning may be broadly distinguished by the agent’s ability to extract information from neighbors whose characteristics (types) differ from his own. When learning is efficient, the agent conditions for differences in his neighbors’ types in learning from their prior decisions and experiences. Consequently, all neighbors have equal influence in the agent’s learning function, regardless of their distance from him in type space. The conditioning described above uses knowledge of the production process which, if incorrect, would result in biased inference about the quality of the new technology. Consequently, the agent may rationally prefer to use relatively inefficient learning rules that discount valuable information from distant (in type space) neighbors, but place little structure on the production function. The implications of the alternative learning models are tested with data from Indian agriculture. The observed pattern of spatial diffusion is found to be consistent with fairly efficient learning in which agents condition, albeit crudely, for differences in their neighbors’ types when learning from them. Information will consequently diffuse quite rapidly, even in populations with heterogeneous characteristics.

Thesis Supervisor: Abhijit V. Banerjee
Title: Associate Professor of Economics
Acknowledgements

I am grateful to my advisor Abhijit Banerjee for his support and guidance throughout the dissertation research. I also thank David Genesove and Franklin Fisher for their comments and advice on the papers that I have written over the past four years. Chapter 2 of this thesis was previously published as "Milk supply behavior in India: Data integration, estimation and implications for dairy development," by K.D. Munshi and K.S. Parikh, in the Journal of Development Economics, Vol.45, No.2 (1994). The article is reprinted in this thesis with the permission of the publishers Elsevier Science B.V..
Chapter 1:

Investment Behavior, Financial Structure and Corporate Control:
The Case of the Japanese Pharmaceutical Industry

1. Introduction

This paper is motivated by a simple stylized fact. In the mid-1980s, Japanese pharmaceutical companies shifted their R&D focus away from antibiotics into new areas, such as the development of anticancer and cardiovascular products. The pharmaceutical industry is regulated by the Ministry of Health and Welfare (MHW) in Japan and drug prices are set by the MHW. The changes in R&D strategy coincided with a period in which the MHW systematically cut all drug reimbursement prices, with the sharpest cuts in antibiotics relative to the other therapeutic classes (Reich, 1990). The price-cuts are a popular explanation in Japan, among bankers, bureaucrats and in the pharmaceutical industry, to account for the shift in intra-firm R&D allocations across therapeutic classes in the 1980s. In contrast, this paper proposes an alternative explanation for these shifts, which is based on financial deregulation that occurred in Japan around the same period.

A well known feature of Japanese corporate finance is the close relationship between the firm and a particular bank, often referred to as the "main bank". Theoretical advances in financial economics addressing information problems in capital markets have motivated a substantial empirical literature in recent years. Hoshi, Kashyap and Scharfstein (1990b, 1991) show quite convincingly for the Japanese case that a close firm-bank relationship allows firms greater access to capital, presumably because information problems are reduced. While the information problems associated with the debt contract have been well articulated in the literature (this goes back to Jensen and Meckling, 1976), reducing these problems by strengthening the firm-bank relationship creates a new set of imperfections. As the bank increases its influence over the firm’s discretionary investments, there will be a systematic trend toward excessively safe projects. This follows from the concavity of the bank’s payoff function under the standard debt contract which

---

1 This chapter is co-authored with Michael R. Reich.
results, in turn, because the bank receives the same return in all success-states. Antibiotics are generally considered to be safer investments, with regard to uncertainty in product development, than anticancer or cardiovascular drugs. Our alternative explanation for the shift in R&D focus in the 1980s thus hypothesizes that Japanese pharmaceutical firms were traditionally forced by their banks into choosing excessively safe R&D projects, with a disproportionate share of total R&D expenditure allocated to antibiotics. Financial deregulation weakened the bank's influence, or aligned its incentives more closely with the firm's, resulting in a shift in R&D allocations towards riskier therapeutic classes.

The specific objective of this paper is to test the alternative explanations for the shift in R&D allocations, that we described above. The results of our empirical exercise have particular policy relevance in light of recent efforts to introduce bank-oriented financial institutions in developing countries and transforming socialist economies. It has been argued that the resurgence of interest in banks as a theoretical and practical alternative to securities markets can be attributed, in large part, to the industrial success of Japan and Germany, both of which relied on bank-financing for their post-war reconstruction and development (Aoki, Patrick and Sheard, 1994). Attempts to replicate such financial institutions in other economies would be ill-founded if the close-firm bank relationship were to distort the firm's investment-portfolio significantly. Such distortions are likely to be most severe for firms having access to high-risk projects, without liquid assets to cover their liabilities. High-tech industries generally fit this characteristic since human capital is typically the dominant input in risky R&D. In that sense, the pharmaceutical industry provides a particularly suitable case study to test for the presence of investment-distortions. since the development of new products is extremely human-capital intensive and associated with high risk. If such distortions are found to be relatively mild in the pharmaceutical industry, then they are unlikely to be significant elsewhere in the economy.

Bank loans have traditionally served as the chief source of external finance in the non-financial corporate sector in Japan and the main bank typically provides the largest share of such capital to the firm. The pharmaceutical industry is an outlier in this regard, however, since bank debt

2 See, for instance, Stiglitz and Weiss, 1981. Success-states refer in this case to states of nature in which the firm is able to repay its debt.
accounts for only a small share of firms’ liabilities in that industry. The main bank nevertheless continues to play an important role in the financial arrangements of pharmaceutical companies, serving as a guarantor or trustee administrator of their other liabilities. The main bank typically assumes responsibility for these liabilities when a firm is in financial distress, even if it is not legally obligated to do so. The role of the main bank is not confined to the provision of financial services. Banks are usually among the largest shareholders in Japanese public corporations, and the main bank is typically the largest shareholder among the banks. The main bank thus operates more as a shareholder with complete liability, than as a lender, in the Japanese pharmaceutical industry. The bank’s payoff function nevertheless continues to remain concave since it receives only a share of the firm’s payoffs in success-states, as a shareholder, but is responsible for all its liabilities in failure-states. The distortion towards excessively safe projects will continue to be obtained when the bank controls investment decisions within the firm.

The main bank’s responsibility for the firm’s liabilities in failure-states distinguishes its objectives from those of the other shareholders in the firm. The bank prefers less debt when investment-risk increases since it is responsible for the firm’s liabilities in failure-states. In contrast, the shareholders prefer more debt when investment-risk increases since they enjoy limited liability under the debt contract. Since the investment portfolio and the financial structure are jointly determined within the firm, we derive implications for the equilibrium correlation between these variables. The antibiotic-share (of total R&D expenditure) is shown to be positively correlated with the debt-equity ratio in equilibrium, under bank control. The sign of the equilibrium correlation is reversed under shareholder control. In practice, of course, neither the bank nor the shareholders will exert complete control over the firm’s investment and financial decisions. Empirical estimates of the correlation between the antibiotic-share and the debt-equity ratio will nevertheless allow us to infer the net outcome of the underlying intra-firm bargaining process between the bank and the shareholders. The estimated sign of the correlation also reveals the direction of investment-distortion. Recall, from our previous discussion, that the firm chooses excessively safe projects under bank control and excessively risky projects under shareholder control.

---

3 For instance, Campbell and Hamao (1994) base their restrictive definition of “main bank firms” on the volume of outstanding bank loans. In their sample of all non-financial firms listed on the Tokyo Stock Exchange, the pharmaceutical industry has the lowest proportion of main bank firms in an economy-wide total of 26 industries.
At a more general level, our econometric strategy provides an alternative solution to the endogeneity problem that plagues empirical work in corporate finance. Since firm-specific variables are typically endogenously determined, the standard approach in the absence of suitable instruments has been to exogenously group firms, or exploit financial-market shocks, when studying the relationship between investment decisions and financial conditions. In contrast, we derive the equilibrium correlation between the investment decision and the financial structure, recognizing that those variables are jointly determined within the firm. For our particular case, the OLS estimate of this correlation then directly reveals the underlying corporate-control regime. This empirical approach clearly has broader applicability in empirical corporate finance.

We obtain R&D expenditures by therapeutic class (antibiotics, cardiovascular, anticancer, other) for eight firms over an eleven year period, 1980-90. The firms in our sample are fairly representative of the total of 34 pharmaceutical firms listed on the first section of the three major Japanese stock exchanges (2 digit code: 45). The antibiotic-share is regressed on the debt-equity ratio to test the implications discussed above. Since the correlation implications are conditional restrictions, holding constant the total R&D expenditure, bank ownership in the firm and relative prices, those variables are also included in the regression equation. The antibiotic-share and the debt-equity ratio are found to be negatively correlated, implying that shareholders exert net control over decision-making within the firm and that the R&D allocation is excessively risky. A close firm-bank relationship does not appear to distort the investment-portfolio toward excessively safe projects in the Japanese pharmaceutical industry. In addition, the change in relative drug-prices appears to be an important determinant of the shift in R&D allocation across therapeutic classes in the sample period, supporting the conventional price-cuts hypothesis.

While the shareholders may exert net control over decision-making within the firm, the bank must continue to retain significant influence over such decisions if the close firm-bank relationship is to successfully reduce information problems. In the Japanese case it is commonly believed that

---


5 The three major stock exchanges are located in Tokyo, Osaka and Nagoya. Large firms tend to be listed on the first section. The eligibility requirements are based on the total number of shares issued, stock distribution, monthly stock turnover and dividends.
the bank ensures that suitable investments are chosen by placing "external" directors and auditors on the boards of its client firms. In addition, it consolidates its position through mid-career transfers of managers to these firms, since many of these managers ultimately rise to the board as "internal" directors (Aoki, Patrick and Sheard, 1994). A testable implication of this corporate governance hypothesis, as a mechanism for reducing information problems between the firm and the bank, is that the underlying corporate control regime should be responsive to exogenous economic and financial shocks.

As an additional empirical exercise, we consequently study the stability of the debt-equity ratio and bank ownership coefficients under alternative economic and financial regimes. Recall here that the debt-equity ratio coefficient reflects underlying corporate control within the firm. A negative coefficient is associated with shareholder control, whereas the sign is reversed under bank control. The bank ownership coefficient is interpreted to reflect the influence exerted by the bank within the firm. In general, we expect this coefficient to increase as the incentives of the bank and the shareholders diverge. Corporate control appears to respond to changes in the economic and financial environment, both over time as well as across firms, providing empirical evidence in favor of the corporate governance hypothesis as the mechanism through which information problems are reduced.

The paper is organized in five sections. Section 2 provides a brief description of the Japanese financial deregulation of the 1980s, especially with regard to how this affected pharmaceutical firms. Section 3 presents a simple model of R&D investment behavior, which permits us to derive the equilibrium correlation between antibiotic-share and the debt-equity ratio. As discussed above, the sign of this correlation identifies the nature of the underlying corporate control within the firm. Section 4 presents the econometric analysis and the main empirical result; the shareholders appear to exert net control over decision-making within the firm despite its close relationship with the main bank. Additional results on the responsiveness of corporate control within the firm, to changes in the economic and financial environment over time and across firms, are also included in this section. Section 5 concludes the paper.
2. The Changing Structure of Japanese Corporate Finance

Japanese firms have traditionally relied on bank loans as their main source of external finance. Bank loans in Japan do not require collateral in most cases. In addition, bank debt served as a regular source of capital, even in bad times, in the high-growth post-war years (Rosenbluth, 1989). As growth slowed down in the 1970s and profits declined, Japanese firms began to explore alternative sources of capital to avoid the costs associated with bank financing. Given the restrictions on domestic bond and equity issuance, many firms began to issue bonds on the Euromarket, where there were no collateral requirements. According to the Ministry of Finance (cited in Hoshi, Kashyap and Scharfstein, 1990a) by 1983, Japanese firms raised almost half their capital in overseas markets.

To respond to the Euro-bond market, the collateral restrictions on domestic straight bonds and convertible bonds were relaxed in 1983. By the end of 1984, 20 companies were permitted to issue straight bonds and 110 could issue convertible bonds on the domestic market. Of the convertible bond-eligible firms, nearly half were permitted to issue unsecured convertible bonds (Rosenbluth, 1989). The domestic bond-issuance criteria were gradually relaxed over the 1980s and by 1987, 180 firms could issue unsecured straight bonds and 330 firms could issue unsecured convertible bonds (Hoshi, Kashyap and Scharfstein, 1990a). Finally, by 1989 these numbers were increased, to approximately 300 and 500 firms, respectively (Campbell and Hamao, 1994).

This process of financial deregulation resulted in a dramatic shift away from bank-financing in Japan. In 1975, more than 90% of the corporate debt of public companies was bank debt; in 1992 it was less than 50% (Hoshi, Kashyap and Scharfstein, 1993). The shift from bank loans to corporate bonds does not, however, necessarily imply a decline in the banks’ responsibility for their firms. Straight and warrant bonds issued on the Euromarket, especially in the early years, required a guarantor since the firms were not very well known (Campbell and Hamao, 1994). The firms’ main bank typically guaranteed such issues making these bonds effectively a securitized version of bank loans. Similarly, secured bonds on the domestic market required a trustee to administer the firms’ collateral. While the trustee bank is not legally obligated to do so, historically it has always bought back the issue in the event of a corporate bond default (Campbell and Hamao, 1994, Aoki, Patrick and Sheard, 1994). In that sense, the role of a trustee bank does not differ from that of a guaranteeing bank. Main banks, in their roles as guarantor
and trustee administrator, would have continued to maintain their influence within their client firms despite the decline in the relative volume of bank loans.

Convertible bonds issued on the Euromarket did not require bank guarantees, even in the early years (Campbell and Hamao, 1994). In addition, a relatively large number of firms were permitted to issue unsecured convertible bonds on the domestic market from 1987 onward. Many Japanese firms, including a number of companies in our sample, took advantage of the stock market boom of the late 1980s to issue equity and equity-related financial instruments, especially convertible bonds. The share of convertible bonds in all bond financing increased from 30% in 1977 to 60% in 1986 (Hoshi, Kashyap and Scharfstein, 1990a). While banks were not legally responsible for such additions to firms' capital stock, they would most likely have covered their client-firms' liabilities in the event of a bond default, following the argument outlined above. Financial deregulation consequently increased the bank's liabilities, at least in the short-run, by allowing firms greater access to public debt. Since convertible bonds are equity-linked, however, positive shocks to the debt-equity ratio following the issue of convertible bonds would have been followed by a decline in debt-equity as such bonds were converted to equity. The late-1980s were consequently characterized by a series of exogenous shocks, both positive as well as negative, to the debt-equity ratio.

3. The Model

Consider a simple model of R&D investment in which the firm consists of an entrepreneur and the bank's representative. The bank provides a share, \( \theta \), of the firm's equity, \( E \). The firm chooses its financial structure (debt-equity ratio, \( R \)) in period-0 and consequently approaches the bank for a loan, \( B \). The bank's loan officer sets the (gross) interest rate on the loan, \( r_B \), in period-1. Finally, in period-2 the firm chooses its R&D portfolio given the debt-equity ratio and interest rate that were chosen in the previous periods. The important point to note here is that the firm operates in two distinct markets, the financial market and the product market, in our model. We will take advantage of this feature of decision-making within the firm later in the econometric section of the paper.

With regard to the production technology, two projects are available to the firm; a risky project and a safe project. The risky project provides payoff, \( P_{II} \), with probability, \( F_{II}(K_{II}) \), when it is
successful, where $K_h$ refers to the amount of capital invested in the project. Similarly, the safe project provides payoff, $P_s$, with probability, $F_s(K_s)$. Both projects provide zero payoff in the failure-state. R&D production exhibits decreasing returns to scale; $F_h'(K_h) > 0$, $F_h''(K_h) < 0$, $F_s'(K_s) > 0$, $F_s''(K_s) < 0$. We assume, $P_h > P_s$, $F_h(K_h) < F_s(K_s)$, $\forall K_h, K_s$ in the range of feasible R&D allocations.

Following the description above, the capital structure is,

$$E + B = K_h + K_s = K.$$ 

We normalize total capital, $K = 1$, which allows us to interpret $K_s$ as the share of total R&D allocated to the safe project. The cost of debt-capital for the bank is also set to unity. The debt-equity ratio, $R = B/E$. Hence, $B = R/(1 + R)$ and $E = 1/(1 + R)$.

Note that all the capital is assumed to be allocated as R&D expenditure. Production costs are negligible when compared with R&D allocations, and product sales, in the pharmaceutical industry and are typically ignored in studies of pharmaceutical investment behavior.

Three cases will be considered below; optimal investment without capital market imperfections, and entrepreneur (shareholder) and bank-representative control when information problems are present. For the latter cases, we will model financial structure and investment behavior as outcomes of a sequential game between the firm and the bank.

**Case 1: Optimal Investment**

Since the information problems considered in our model are associated with the debt contract, we consider the case in which all the capital is raised through equity financing. The firm's problem can then be expressed as,

$$\max_{K_h} \left( P_h + P_s F_h(K_h) F_s(K_s) + P_h F_h(K_h)(1 - F_s(K_s)) + P_s F_s(K_s)(1 - F_h(K_h)) - C_E(K) \right)$$

(1)

where, $C_E(\cdot)$ is the cost of raising equity. $C_E'(\cdot) > 0$, $C_E''(\cdot) > 0$.

The firm's problem can be simplified as,

$$\max_{K_h} \left( P_h F_h(K_h) + P_s F_s(K_s) - C_E(K) \right)$$

(1')

The first-order condition is obtained as,
\[ \frac{F_L'(K_L^*)}{F_H'(K_H^*)} = \frac{P_H}{P_L} \]  

where, \( K_L^* \), \( K_H^* \), represent first-best R&D allocations to the safe and risky projects, respectively.

**Case 2: Entrepreneur control**

We solve the sequential game between the firm and the bank by backward induction. The entrepreneur takes the debt-equity ratio, \( R \), and the interest rate, \( r_b \), as given in period-2 when choosing his R&D allocation, \( K_L(r_b, R) \). Working forward, the bank’s loan-officer takes, \( R \), as given when fixing the interest rate, \( r_b(R) \), in period-1. We consequently obtain, \( K_L(R) \). Finally, the firm chooses, \( R \), in period-0 recognizing its future effect on \( r_b \) and \( K_L \).

**Period-2: Product market decision**

The entrepreneur’s problem, taking \( r_b \) and \( R \) as given, may be expressed as,

\[
\max_{K_L} (1-\theta)(P_H + P_L - r_bB)(F_H(K_H)F_L(K_L) + (P_H - r_bB)F_H(K_H)(1 - F_L(K_L)) + (P_L - r_bB)F_L(K_L)(1 - F_H(K_H))) - C_L((1-\theta)E)
\]

The equation above can be simplified as,

\[
\max_{K_L} (1-\theta)(P_HF_H(K_H) + P_LF_L(K_L)) - (1-\theta)r_bBG(K_L) - C_L((1-\theta)E)
\]

where, \( G(K_L) = [F_H(K_H) + F_L(K_L) - F_H(K_H)F_L(K_L)] \) is the probability that the loan is repaid. We assume that the firm is able to repay its loan if the low-risk project succeeds, \( P_L > r_bB \).

The first-order condition is obtained as,

\[
-P_HF_H'(K_H) + P_LF_L'(K_L) - r_bBG'(K_L) = 0
\]

Substituting for \( G'(K_L) \), equation (4) may be written as,

\[
\frac{F_L'(K_L^*)}{F_H'(K_H^*)} = \frac{P_H - X_H}{P_L - X_L}
\]

where, \( K_L^{**} \), \( K_H^{**} \) are optimal R&D allocations under entrepreneur control.

\[ X_H = r_bB(1 - F_L(K_L)) \quad X_L = r_bB(1 - F_H(K_H)) \]
Comparing equation (2) and equation (4'), $K_L^{**} < K_L^*$, since $X_H < X_L$. Investment is consequently distorted toward excessively risky projects, $\forall \ r_b > 0$.

The R&D allocation is obtained from equation (4) as a function of $r_b$ and $R$, $K_L(r_b, R)$. The following comparative statics results are obtained,

$$\frac{dK_L}{dr_b} \frac{BG'(K_L)}{S_1} < 0 \quad (5)$$

$$\frac{dK_L}{dR} \frac{r_b^G(K_L)}{(1+R)^2S_1} < 0 \quad (6)$$

where, $S_1 < 0$ is the second-order condition from equation (4).

$G'(K_L) = F_L'(K_L)(1-F_H(K_H)) - F_H'(K_H)(1-F_L(K_L)) > 0$, since $F_L'(K_L)/F_H'(K_H) > P_H/P_L > 1$, from equation (4').

**Period-1: Setting the interest rate**

The bank’s loan-officer sets the interest rate, $r_b$, after observing the firm’s demand for credit, $B$, since the firm’s debt-equity ratio $[B = R/(1+R)]$ reveals its investment portfolio in period-2. To begin with, we make the somewhat unrealistic assumption that the loan officer sets $r_b$ to break even, $r_bBG(K_L) - B = 0$.

$$r_b = \frac{1}{G(K_L)} \quad (7)$$

From equation (4), $K_L(r_b, R)$. Substituting in equation (7), we obtain, $r_b(R)$ and hence, $K_L(R)$. The comparative statics result is,

$$\frac{dr_b}{dR} = \frac{dr_b}{dK_L} \cdot \frac{dK_L}{dR} = -\frac{G'(K_L)}{[G(K_L)]^2} \frac{dK_L}{dR} > 0 \quad (8)$$

since $dK_L/dR < 0$ from equation (6).

In reality, the loan-officer will account for the bank’s equity-share in the firm when setting the interest rate. Nevertheless, as long as the bank is responsible for all the firm’s liabilities in failure states, its incentives will not be aligned with those of the shareholders. The firm under entrepreneur control will always choose an investment portfolio that is too risky, and demand too
much debt, relative to the bank’s preferred levels for these decision-variables. The loan-officer will use the interest rate to align the firm’s decisions, under entrepreneur control, more closely with the bank’s incentives. The comparative statics result, $\frac{dr}{dR} > 0$, is consequently quite general since by increasing the interest rate the bank is able to reduce the demand for debt.

**Period-0: Financial market decision**

The entrepreneur chooses his financial structure, $R$, by recognizing the effect that it will have on the interest rate, $r_b$, and his investment decision, $K_L$, in subsequent periods. The entrepreneur’s problem is,

$$\max_R \quad (1-\theta)[P_H F_H(K_H) + P_L F_L(K_L)] - (1-\theta)r_bBG(K_L) - C_E((1-\theta)E)$$

(9)

Noting that $r_b(R), K_L(R)$, the first-order condition is obtained as,

$$\left[-P_H F_H'(K_H) + P_L F_L'(K_L) - r_bBG'(K_L)\right]\frac{dK_L}{dR} - BG(K_L)\frac{dr_b}{dR} - \frac{r_bG(K_L)}{(1+R)^2} + \frac{C_E((1-\theta)E)}{(1+R)^2} = 0$$

(10)

Substituting from equation (4),

$$-BG(K_L)\frac{dr_b}{dR} - \frac{r_bG(K_L)}{(1+R)^2} + \frac{C_E((1-\theta)E)}{(1+R)^2} = 0$$

(10')

We earlier obtained, $dK_L/dR<0$ from equation (6). Since $K_L$ and $R$ are jointly determined, we obtain the symmetric condition from equation (10'). The firm chooses, $R$, in the financial market (in period-0) with a target $K_L$ in mind, which is realized in equilibrium.

$$\frac{dR}{dK_L} = \frac{BG'(K_L)}{S_2} \frac{r_bG'(K_L)}{(1+R)^2}$$

(11)

since $dr_b/dR > 0$ from equation (8) and $S_2 < 0$ is the second-order condition from equation (10').

The product-market/financial-market equilibrium derived above is presented graphically in Figure 1(a). The important point to note here is that both $K_L-R$ curves are negatively sloped; $dK_L/dR < 0$, $dR/dK_L < 0$. The equilibrium $K_L$ and $R$ are obtained as the intersection of the $K_L-R$ curves.
Case 3: Bank-representative control

The Nash equilibrium solution is derived, as in the previous case, by backward induction. As discussed earlier, the bank operates more like a shareholder with complete liability, than as a creditor, in the Japanese pharmaceutical industry.

Period-2: Product market decision

The bank-representative’s problem is,

$$\max_{\kappa_t} \theta\left[(P_H + P_L - r_B B)F_H(K_H)F_L(K_L) + (P_H - r_B B)F_H(K_H)(1 - F_L(K_L)) + (P_L - r_B B)F_L(K_L)(1 - F_H(K_H))\right] + r_B B G(K_L) - B - C_g(\theta E)$$

The above equation can be simplified as,

$$\max_{\kappa_t} \theta\left[(P_H F_H(K_H) + P_L F_L(K_L)) + (1 - \theta) r_B B G(K_L) - B - C_g(\theta E)\right]$$

(12)

The first-order condition is obtained as,

$$\theta\left[-P_H F_H'(K_H) + P_L F_L'(K_L)\right] + (1 - \theta) r_B B G'(K_L) = 0$$

(13)

Substituting for $G'(K_L)$, equation (13) may be written as,

$$\frac{F_L'(K_L^{**})}{F_H'(K_H^{**})} = \frac{P_H + Y_H}{P_L + Y_L}$$

(13')

where, $K_L^{**}$, $K_H^{**}$ are optimal R&D allocations under bank control.

$$Y_H = [(1 - \theta) / \theta] r_B B (1 - F_L(K_L)) \quad Y_L = [(1 - \theta) / \theta] r_B B (1 - F_H(K_H))$$

Comparing equation (2) and equation (13'), $K_L^{**} > K_L^{*}$, since $Y_H < Y_L$. Investment is distorted toward excessively safe projects, $\forall r_B > 0$.

As before, the R&D allocation is obtained from equation (13) as, $K_L(r_B, R)$. The comparative statics results are as follows,
\[
\frac{dK_L}{d\theta} = - \left( \frac{1 - \theta}{\theta} \right) \frac{BG'(K_L)}{S_3} > 0
\] (14)

\[
\frac{dK_L}{dR} = - \left( \frac{1 - \theta}{\theta} \right) \frac{r_b BG'(K_L)}{(1 + R)^2 S_3} > 0
\] (15)

where, \( S_3 < 0 \) is the second-order condition from equation (13) and \( G'(K_L) > 0 \) from equation (13').

**Period-1: Setting the interest rate**

Since investment and financial decisions within the firm are now controlled by the bank’s representative, the firm’s and the loan-officer’s incentives are aligned. Recall here that the bank’s representative behaves like a shareholder with complete liability in this case. The loan-officer consequently no longer sets the interest rate, \( r_b \), to influence the firm’s investment decisions. Instead, it will pass loans through to the firm at-cost. The interest rate, \( r_b \), is fixed in this case (equal to unity) and \( dr_b/dR = 0 \).

**Period-0: Financial market decision**

The bank-representative’s problem is,

\[
\max_{r} \quad \theta [P_H F_H(K_L) + P_L F_L(K_L)] + (1 - \theta) r_b BG(K_L) - B - C_E(\theta E)
\] (17)

The first-order condition is obtained as,

\[
\left[ \theta (-P_H F_H'(K_L) + P_L F_L'(K_L)) + (1 - \theta) r_b BG'(K_L) \right] \frac{dK_L}{dR} + (1 - \theta) r_b G(K_L) \frac{dK_L}{(1 + R)^2} - \frac{1}{(1 + R)^2} + \frac{\theta C_E'(\theta E)}{(1 + R)^2} = 0
\] (18)

Substituting from equation (13),

\[
(1 - \theta) r_b G(K_L) - 1 + \theta C_E'(\theta E) = 0
\] (18')

As discussed earlier, the firm chooses \( R \) in the financial market (in period-0) with a target \( K_L \) in mind, which is realized in equilibrium. From equation (18'),

17
\[
\frac{dR}{dK_L} = \frac{-(1-\theta)\theta G'(K_L)}{S_i} > 0
\] (19)

where, \( S_i < 0 \) is the second-order condition from equation (18').

The product-market/financial-market equilibrium derived above is presented graphically in Figure 1(b). The equilibrium \( K_L \) and \( R \) are derived as the intersection of the \( K_L - R \) curves. The important point to note here is that both \( K_L - R \) curves are positively sloped; \( dK_L/dR > 0, dR/dK_L > 0 \). We will take advantage of this equilibrium property when deriving the sign of the equilibrium correlation between the antibiotic-share \( (K_i) \) and the debt-equity ratio \( (R) \), in the next section.

4. Estimation

Recall the implications that we derived in the previous section:
under entrepreneur control - \( dK_L/dR < 0, dR/dK_L < 0 \)
under bank control - \( dK_L/dR > 0, dR/dK_L > 0 \).

The intersection of the \( K_L - R \) schedules determines the equilibrium, \( K_L \) and \( R \), in each case. The data that we obtain in practice, are equilibrium realizations of these variables. This section consequently begins with the derivation of the equilibrium correlation between the antibiotic-share \( (K_i) \) and the debt-equity ratio \( (R) \), under alternative entrepreneur and bank control regimes. We then proceed to describe the data and discuss the results of the estimation. We conclude this section with an extension of the empirical analysis to study how the \( K_L - R \) correlation responds to changes in the economic and financial environment, over time as well as across firms.

4.1 Equilibrium \( K_L - R \) correlation

The implications derived above may be expressed, with the addition of stochastic terms, as follows,
\[(K_L)_i = \alpha R_i + \epsilon_i^1 \]  \hspace{1cm} (a)

\[R_i = \beta (K_L)_i + \epsilon_i^2 \]  \hspace{1cm} (b)

where, equation (a) represents the product-market decision and \(\epsilon_i^1\) are product-market shocks. Similarly, equation (b) represents the financial-market decision with corresponding shocks, \(\epsilon_i^2\).\(^6\)

Following the implications derived above,
under entrepreneur control - \(\alpha < 0, \beta < 0\)
under bank control - \(\alpha > 0, \beta > 0\).

Without suitable instruments, neither equation (a) nor equation (b) can be identified and, in practice, the equilibrium correlation will be estimated using OLS on equation (a).

\[
\hat{\alpha} = \alpha + \frac{\beta}{1 - \alpha \beta} \frac{1}{T} \sum_i (\epsilon_i^1)^2 + \frac{1}{1 - \alpha \beta} \frac{1}{T} \sum_i \epsilon_i^1 \epsilon_i^2 \]

\[\frac{1}{T} \sum_i R_i^2 \]

Assuming that product-market and financial-market shocks are uncorrelated, \(E(\epsilon_i^1 | \epsilon_i^2) = 0\),

\[
\text{plim} \hat{\alpha} = \pi = \alpha + \beta \frac{\sigma_i^2}{\sigma_R^2} \]

The usual equilibrium stability condition requires, \(\alpha \beta < 1\). Consequently, the sign of the equilibrium correlation, \(\pi\), still reveals the nature of the underlying corporate control.
under entrepreneur control - \(\alpha < 0, \beta < 0 \Rightarrow \pi < 0\)
under bank control - \(\alpha > 0, \beta > 0 \Rightarrow \pi > 0\).

The correlation derived above is, in fact, a conditional restriction. Recall that the comparative statics results derived in the previous section were obtained for a given bank share (\(\theta\)), relative prices (\(P_u/P_L\)), and total R&D expenditure (\(K\), which was normalized to unity). Equation (a) will consequently be extended, when estimating the model, to include these additional variables.

\(^6\) \(\epsilon_i^1\) and \(\epsilon_i^2\) may be interpreted as determinants of the antibiotic-share and the debt-equity ratio that are unobserved by the econometrician, but unrelated to the joint choice of \(K_L\) and \(R\). For instance, \(\epsilon_i^1\) could represent firm-specific technological shocks that affect the choice of the antibiotic-share in a given year, and are known to both the firm and the bank.
4.2 The data

As described earlier, we estimate the model using R&D expenditure data for eight firms over an eleven year period (1980-90). We obtain R&D expenditure shares for antibiotics, cardiovascular and anticancer therapeutic classes. The financial deregulation explanation for the shift in R&D allocations away from antibiotics in the 1980s hypothesizes that pharmaceutical firms were traditionally constrained by their banks into choosing excessively safe projects (R&D in antibiotics). While we stated earlier that antibiotics are generally considered to be the safest investment, no justification was presented to support that claim.

With regard to uncertainty in product development, technology favors antibiotic research. Systematic screening for antimicrobial activity provides an effective tool in the discovery of antibiotic products, and fermentation technology for antibiotic drug development is well understood and well developed in Japan (Takeuchi, 1987). This reduces the uncertainty of antibiotic drug development significantly. Cardiovascular drugs, in contrast, include powerful agents whose mechanism of action is well known, as well as less effective agents whose mechanism of action is often unclear (Cockburn and Henderson, 1993). Finally, anticancer drugs would also be considered as a risky therapeutic class since few products with proven efficacy have been discovered in that category. A survey on the scientific uncertainty of drug development across different therapeutic classes, conducted among 283 researchers in 17 Japanese firms and 47 researchers in 20 U.S. firms supports our ranking, showing antibiotics to have a low level of scientific uncertainty, and anticancer to be high, with cardiovascular to be moderate (Innami, 1992). Since drug prices are set by the MHW in Japan, scientific and economic uncertainty are likely to be highly correlated.

Our model also requires data on relative sales per product, as a determinant of the distribution of R&D expenditures. Drug prices are set by the Ministry of Health and Welfare (MHW) as a single nationwide reimbursement rate under the universal health insurance system. Physicians prescribe as well as dispense drugs; only about 12% of the total number of prescriptions is dispensed through pharmacies. To encourage utilization of their products, pharmaceutical

---

7 We were unable to obtain R&D expenditures over the entire estimation period for three firms in our sample. Data for these firms begin in 1981, 1984 and 1986, respectively. We thus have a total of 77 observations for the estimation procedure.
companies (or their wholesalers) typically sell drugs to physicians at a discount (averaging around 24%) below the set reimbursement rate. The "doctor's margin" or "yakka saeki" accounts for a substantial fraction of total health care expenditures in Japan. Total expenditure on pharmaceutical products represented a major portion of total medical costs in that country, declining from 22% in 1980 to 17% in 1986. This contrasts with the corresponding figure of 7% for the U.S. throughout the 1980s (OECD, 1990).

In the 1980s, fiscal pressures prompted the MHW to slash drug reimbursement prices in an effort to reduce the "doctor's margin," with an average cumulative reduction in drug prices of nearly 50% for the decade. The severity of these price cuts varied by therapeutic class; for instance, antibiotics were most seriously affected. As discussed earlier, the change in relative drug prices has been proposed as an explanation for the shift in the distribution of pharmaceutical R&D expenditures in the health policy literature (Reich, 1990). Government officials and corporate executives in Japan tend to follow this line of reasoning as well. An additional explanation for the shift away from antibiotics is demand-driven. In the epidemiological transition, as a country grows more wealthy and the population ages, the proportion of mortality from noncommunicable diseases (such as cancer and cardiovascular disease) typically rises, while the corresponding proportion from communicable diseases declines. Japan has followed this pattern as well (Omran, 1971).

The regression equation consequently includes sales (revenue) figures, rather than product prices, to control for changes in prices, as well as in demand, across therapeutic classes. These data were obtained from a private pharmaceutical company data-base allowing us to compute the average sales for a product in each of the therapeutic classes (antibiotics, cardiovascular, anticancer) for each year in the estimation period. We computed averages for the top 10 and top 15 products since sales drop sharply thereafter.

We present descriptive statistics in Table 1, for the variables included in the regression equation. The Table is divided into two sections; the first section reports means and standard deviations for these variables by year, whereas the second section reports the corresponding statistics by firm.
Beginning with the description by year, Table 1 reveals a systematic shift away from antibiotics, as a share of total R&D expenditure, over the 1980s. While we do not report the corresponding figures for anticancer and cardiovascular R&D, shares for these therapeutic classes increased systematically over the decade. This period also saw a steady decline in the sales per product for antibiotics, relative to the other two therapeutic classes, especially in the mid-1980s. To preserve space we present relative sales figures for anticancer-antibiotics and cardiovascular-antibiotics, using data for the top ten selling-products in each year.\(^8\) The corresponding trends, with sales for the top fifteen products, are qualitatively similar.

Main bank ownership declined during the 1980s as well, partly as a response to the Revised Anti-Monopoly Act of 1977 which reduced the bank-ownership ceiling from 10\% to 5\% (effective in 1987). While this certainly provides an exogenous component to the decline in bank ownership over the sample period, the time-lag between the announcement of the Act and its implementation allowed firms some flexibility in determining their bank-ownership levels. As we discussed earlier, the firms in our sample issued convertible bonds in the Euromarket in the early 1980s, as well as in the domestic market after 1987 once the eligibility criteria for unsecured bonds were relaxed. Since convertible bonds are equity-linked, their issue is associated in the long-run with a dilution of the bank’s equity-share in the firm. In addition, public stock-offerings by firms in the sample over this period may also have lead to a reduction in bank ownership.

With regard to the remaining variables included in the regression equation, total R&D expenditures increased systematically over the sample-period. This is most likely an outcome of the increased access to external capital that resulted from the financial deregulation of the 1980s. Finally, no apparent trend in the debt-equity ratio is revealed in Table 1, although debt-equity appears to be generally higher in the first half of the decade. The absence of a time-trend is perhaps fortunate, since it reduces the potential for spurious estimated correlation between the antibiotic-share and the debt-equity ratio, with debt-equity proxying for omitted time-trending variables.

\(^8\) The relative sales per product listed in year, \(t\), in Table 1 actually refers to the realized sales in year, \(t+1\). The assumption here is that a product developed in year, \(t\), reaches the market at the beginning of year, \(t+1\). The subsequent year’s price is consequently the variable of interest when the firm chooses its R&D allocation across therapeutic classes.
Turning to descriptive statistics by firm, Table 1 reveals considerable variation across firms for the variables of interest. The antibiotic-share ranges from as little as 7%, to as much as 33%. Significant variation is also observed in the total R&D expenditures across firms, as well as in the debt-equity ratio. It is interesting to note that firm 6, with the lowest antibiotic-share among the firms in our sample, displays the highest debt-equity ratio. In contrast, firm 3, with the highest antibiotic-share, is associated with the second-lowest debt-equity ratio. This negative correlation between the antibiotic-share and the debt-equity ratio will be obtained more generally as a result of the regression estimation.

4.3 Estimation results

As discussed previously, the antibiotic-share is regressed on the debt-equity ratio, relative sales per product (anticancer-antibiotics or cardiovascular-antibiotics), bank-ownership and total R&D expenditures. Firm fixed-effects and year dummies (or time trends) are also included in the regression equation. The sign of the estimated debt-equity ratio coefficient reveals the nature of underlying corporate control. Recall that with entrepreneur (shareholder) control the coefficient is negative, whereas the sign is reversed under bank control.

Inspection of Table 2 reveals a negative coefficient on the debt-equity ratio variable across all the alternative specifications. The estimated coefficient is also statistically significant in all cases, with the exception of the linear time-trend specification. This is perhaps the most robust empirical result of the paper, allowing us to infer with a fair degree of confidence that the firm operates under net shareholder control in the Japanese pharmaceutical industry. As discussed previously, this result has important policy implications since it suggests that a close firm-bank relationship does not necessarily constrain the firm into choosing excessively safe projects.

Turning to the relative sales per product, the coefficient associated with anticancer-antibiotic sales (average of top ten products) is found to be negative as predicted by the price-cuts hypothesis, but is not very precisely estimated, when a linear time-trend is included in the regression equation. A similar result is obtained when anticancer-antibiotic sales are replaced by cardiovascular-antibiotic sales in the regression equation. Since the sales per product does not vary across firms, we are sensitive to the possibility that it simply proxies for unobserved time-trending variables that are omitted from the regression equation, such as changes in R&D
technology. We consequently experiment with non-linear time-trends, including non-linear price-effects as well, in the regression equation. Statistically significant price-effects are now obtained, providing support for the price-cuts hypothesis as an explanation for the shift out of antibiotics in the 1980s.\(^9\) We nevertheless include a final specification in Table 1, which replaces the sales and time-trend variables with year fixed-effects. Such year dummies account for price-effects, changes in R&D technology and other time-effects that are omitted from the regression equation.

The estimated bank-ownership and total R&D expenditure coefficients are both positive in Table 2. Following the previous discussion, these variables are assumed to be affected by shocks in the financial market, but to be orthogonal to product-market shocks. Their coefficient estimates are nevertheless biased due to the endogeneity of the debt-equity ratio variable. To see the effect of the endogeneity, consider the multivariate extension to equation (a) and equation (b).

\[
(K_t)_i = \alpha R_t + \gamma X_t + \epsilon_i^1 \\
R_t = \beta (K_t)_i + \epsilon_i^2
\]

\[
\hat{\alpha} = \frac{\left[ \frac{1}{T} \sum X_t^2 \right]^2 - \left[ \frac{1}{T} \sum (K_t)_i R_t \right]^2}{\left[ \frac{1}{T} \sum R_t^2 \right] \left[ \frac{1}{T} \sum X_t^2 \right] - \left[ \frac{1}{T} \sum R_t X_t \right]^2}
\]

Assuming \(E(\epsilon_i^1 \epsilon_i^2) = 0\), \(E(X_i \epsilon_i^1) = 0\),

\[
\text{plim} \hat{\alpha} = \alpha + \left( \frac{\beta}{1 - \alpha \beta} \right) \left[ \frac{\sigma_{\epsilon^1}^2}{\sigma_R^2} \left[ \frac{1}{1 - (\sigma_{R_t}^2)^2/\sigma^2_{\epsilon^1}} \right] \right]
\]

Comparing equation (21) and equation (23), the bias term, \(\text{plim} \hat{\alpha} - \alpha\), continues to be negative while increasing in absolute magnitude, since \(\sigma_R^2 \sigma_{\epsilon^1}^2 - (\sigma_{R_t})^2 > 0\). The equilibrium implications derived earlier consequently are unaffected by the inclusion of additional variables in the regression equation.

---

\(^9\) In contrast with the results described above, extremely weak price effects are obtained when the average sales for the top ten products is replaced by the average sales for the top fifteen products. This could imply that firms employ a fairly high cut-off when computing the expected sales that they would expect if a product were successfully developed.
\[
\hat{\gamma} = \left(\frac{1}{T} \sum R_i^2\right)^{-1} \left(\frac{1}{T} \sum (K_i)X_i - \frac{1}{T} \sum RX_i\right) \left(\frac{1}{T} \sum (K_i)R_i\right) - \left(\frac{1}{T} \sum X_i^2\right) - \left(\frac{1}{T} \sum RX_i\right)^2
\]

Assuming \(E(\varepsilon_i^1 \varepsilon_i^2) = 0\), \(E(X_i \varepsilon_i^1) = 0\), as before,

\[
\text{plim} \hat{\gamma} = \gamma - \left[\frac{\beta}{1 - \alpha \beta} \right] \left[\frac{\sigma^2}{\sigma_R^2}\right] \left[\frac{1}{1 - (\sigma_{R_x}/\sigma^2)}\right] \left[\frac{\sigma_{R_x}}{\sigma^2}\right]
\]

The regression results and sample-correlations reveal, \(\alpha < 0\), \(\beta < 0\), \(\sigma_{R_x} > 0\), in our case. The positive coefficient estimates on the bank-ownership and total R&D variables could consequently follow from the bias term in equation (25). Since we are not especially interested in interpreting the total R&D coefficient, that variable will be ignored in the subsequent discussion. The ownership variable is, however, of particular interest to us. In the context of our model, this variables has either an influence or an incentive interpretation. The influence interpretation follows because the bank's representative is presumably able to exert greater control within the firm when the bank's ownership share increases. However, such an increase in ownership will also serve to align the bank's incentives more closely with those of the shareholders. In terms of equation (25), the influence interpretation implies \(\gamma > 0\), whereas the incentive interpretation implies \(\gamma < 0\).

To distinguish between the alternative interpretations for the bank ownership variable, we compare the estimated debt-equity ratio and bank ownership coefficients in Table 2 with the corresponding expressions derived in equations (23) and (25). Under the incentive interpretation, the estimated bank ownership coefficient is positive on account of the bias term in equation (25). Notice that this bias term differs from the corresponding bias term in equation (23) by the factor, \((\sigma_{R_x}/\sigma^2)\). The sample-statistic for \((\sigma_{R_x}/\sigma^2)\) is obtained in our case as 0.878. The bias in equation (25) is consequently smaller (in absolute magnitude) than the bias in equation (23). This implies, in turn, that the absolute value of the estimated debt-equity ratio coefficient (from equation (23)) must be larger than the estimated ownership coefficient (from equation (25)), under the incentive interpretation (with \(\gamma < 0\)). Inspection of the estimated coefficients in Table 2, however, reveals that the ownership coefficient is significantly larger (in absolute value) than the debt-equity ratio coefficient, across all the alternative specifications, rejecting the incentive interpretation. The
influence interpretation for the ownership variable is consequently assumed in the analysis that follows.

4.4 Stability of the debt-equity and bank ownership coefficients

While the shareholders may exert net control over decision-making within the firm, the bank must nevertheless continue to retain significant influence over such decisions if the close firm-bank relationship is to successfully reduce information problems. A testable implication of the standard corporate governance hypothesis, as a mechanism for reducing information problems between the firm and the bank, is that the underlying corporate control regime should be responsive to changes in the economic and financial regime.

As an additional empirical exercise, we consequently study the stability of the debt-equity ratio and bank ownership coefficients, over time as well as across firms. Recall here that the debt-equity ratio coefficient reflects underlying corporate control within the firm. A negative coefficient is associated with shareholder control, whereas the sign is reversed under bank control. As discussed above, the bank ownership coefficient is interpreted to reflect the influence exerted by the bank within the firm. In general, we would expect this coefficient to increase as the incentives of the bank and the shareholders diverge.

To evaluate the stability of the debt-equity ratio and bank ownership coefficients over time, the estimation period is divided into three regimes, with cut-offs in 1983 and 1987. The first cut-off corresponds to a point just prior to the initiation of the price cuts in 1984. The second cut-off corresponds to the year in which the Anti-Monopoly Act, limiting bank ownership, took effect. This is also the year in which convertible-bond eligibility requirements were relaxed in the domestic market and a number of firms in our sample actually issued convertible bonds after this point. The three regimes may be consequently characterized as follows,

regime 1 (1980-82): high (relative) antibiotic prices, financial regulation
regime 2 (1983-86): low (relative) antibiotic prices, financial regulation

A decline in the relative price (sales per product) of antibiotics leads to a shift away from that therapeutic class, both under entrepreneur control and under bank control. The incentives of the
shareholders and the bank are increasingly aligned towards riskier projects, implying a reduction in the debt-equity ratio coefficient.\textsuperscript{10} We cannot, however, determine \textit{a priori} whether incentives within the firm converge in this case, preventing us from predicting the effect on the bank ownership coefficient. Comparison of the corresponding coefficient estimates from regime 1 to regime 2, in Table 3, reveals a decline in the debt-equity ratio coefficient, as predicted. The bank-ownership coefficient also declines, implying convergence in the incentives of the shareholders and the bank. The absence of a decline in the debt-equity coefficient with the OWN/R specification (allowing both R and OWN coefficients to vary over time) is somewhat surprising. The ownership coefficient nevertheless continues to decline in this case.

Turning next to the effect of financial deregulation on corporate control, the exogenous shocks to the debt-equity ratio associated with financial deregulation cause the incentives of the shareholders and the bank to diverge.\textsuperscript{11} The net outcome of the underlying intra-firm bargaining process consequently cannot be predicted \textit{a priori}. It depends, in general, on how incentives as well as influence change within the firm as a response to the financial deregulation. The model does, however, predict that the ownership coefficient will increase as incentives diverge. Comparison of the corresponding coefficient estimates from regime 2 to regime 3, in Table 3, reveals an increase in both the debt-equity ratio coefficient as well as the bank ownership coefficient.

\textsuperscript{10} In terms of the model, an increase in \( P_d/P_l \) leads to a decrease in \( K_l \) under both entrepreneur control and bank control. Note that the distortion in investment under entrepreneur control follows from the, \( r_r B'G'(K_l) \), term in equation (4). \( dK_l/d\theta \) is increasing (in absolute magnitude) in this term. A decrease in \( K_l \) leads to an increase in \( R \) and, hence, an increase in \( r_r \), under entrepreneur control. Ignoring shifts in \( G'(K_l) \), the debt-equity ratio coefficient will consequently increase (in absolute magnitude) when \( P_d/P_l \) increases. By a similar argument, the distortion in investment under bank control follows from the, \( (1-\theta)r_r B'G'(K_l) \), term in equation (13). A decrease in \( K_l \) under bank control leads to a decrease in \( R \), with no effect on \( r_r \). Ignoring shifts in \( G'(K_l) \), as before, the distortion in investment diminishes as \( K_l \) declines. The increase in \( P_d/P_l \) is consequently associated with a decrease in the debt-equity ratio coefficient, which is positive under bank control.

\textsuperscript{11} In terms of the model, financial deregulation is associated with exogenous shocks to the debt-equity ratio, \( R \). Consider a positive shock to \( R \). Under entrepreneur control, an increase in \( R \) is associated with a decrease in \( K_l \) and an increase in \( r_r \). Ignoring shifts in \( G'(K_l) \), the distortion-term in equation (4), \( r_r B'G'(K_l) \), increases with a corresponding increase (in absolute magnitude) in the debt-equity ratio coefficient. Under bank control, the distortion-term in equation (13), \( (1-\theta)r_r B'G'(K_l) \), increases with an increase in \( R \), since \( r_r \) remains constant and shifts in \( G'(K_l) \) are ignored. The debt-equity ratio coefficient consequently increases under bank control. The incentives of the shareholders and the bank diverge in this case. The same result is obtained, following a similar argument, with a negative shock to the debt-equity ratio.
As a final exercise, we explore changes in the corporate control regime across firms. Firms are grouped on the basis of their average debt-equity ratio, taken over the entire sample-period. The bank ownership coefficient is found to decline for firms with large average debt-equity ratios in Table 3. This is an apparently counter-intuitive result, since incentives between the shareholders and the bank should diverge as the debt-equity ratio increases, from the argument above. This follows essentially because the antibiotic-share is increasing in the debt-equity ratio under bank control, whereas it is decreasing under shareholder control. It should be noted, however, that we previously studied the response of underlying corporate control to exogenous shocks in the economic and financial environment, over time. In contrast, the debt-equity ratio is endogenously determined within the firm and its observed level is determined, at least in part, by the bank’s influence within the firm. High debt-equity ratios may consequently follow from low bank influence. Following this argument, we would expect firms with high debt-equity ratios to display a correlation between the antibiotic-share and the debt-equity ratio that more closely approximates the shareholders’ preferences. We obtain weak statistical evidence consistent with this prediction in Table 3. The results of this final empirical exercise provide additional support for the hypothesis that corporate control is responsive to changes in the economic and financial environment. The relatively weak statistical evidence in this case, however, underscores the problems associated with the use of endogenous firm-specific variables to group firms, as a basis for the comparison of corporate control regimes across firms.

5. Conclusion

Two hypotheses are proposed in this paper to explain the observed shift out of antibiotics in the Japanese pharmaceutical industry over the 1980s. The conventional view is that price-cuts initiated by the Ministry of Health and Welfare (MHW) played an important role in this shift. We propose an alternative hypothesis in this paper, which is based on the financial deregulation that occurred around the same period in Japan.

A well known feature of Japanese corporate finance is the close relationship between the firm and its main bank. While the close firm-bank relationship may reduce information problems associated with the debt-contract, it creates a new set of imperfections by allowing the bank greater influence over the firm’s discretionary investments. Such imperfections arise from the concavity of the bank’s payoff function under the standard debt contract, since the bank receives the same
return in all success-states, leading it to favor excessively safe projects. Antibiotics are generally considered to be relatively safe investments, with regard to uncertainty in product development. An alternative explanation for the traditional focus on R&D in antibiotics is that Japanese pharmaceutical companies were constrained by their banks into allocating a disproportionate share of their R&D expenditures to relatively safe antibiotic research. Financial deregulation then may have weakened the bank’s influence, or aligned its incentives more closely with those of the firm, leading to a shift in R&D allocation toward riskier therapeutic classes.

The main bank’s responsibility for the firm’s liabilities in failure-states distinguishes its objectives from those of the other shareholders in the firm. The bank prefers less debt when its investment-risk increases since it is responsible for the firm’s liabilities in failure-states. In contrast, the shareholders prefer more debt when investment-risk increases since they enjoy limited liability under the debt-contract. Since the investment-portfolio and the financial structure are jointly determined within the firm, we derive implications for the equilibrium correlation between these variables. The antibiotic-share and the debt-equity ratio are shown to be positively correlated in equilibrium, under bank control. The sign of the equilibrium correlation is reversed under shareholder control. Recall, from the previous discussion, that the firm chooses excessively safe projects under bank control and excessively risky projects under shareholder control.

We obtain R&D expenditures by therapeutic class for eight firms over an eleven year period. The antibiotic-share is regressed on the debt-equity ratio to test the implications discussed above. Since the correlation implications are conditional restrictions, holding constant the total R&D expenditure, bank ownership in the firm and relative prices, these variables are also included in the regression equation. We obtain strong statistical evidence that shareholders exert net control over decision-making within the firm. A close firm-bank relationship does not appear to distort the investment-portfolio towards excessively safe projects in the Japanese pharmaceutical industry. In addition, the change in relative drug prices appears to be an important determinant of the shift in R&D allocation across therapeutic classes in the sample-period, supporting the conventional price-cuts hypothesis.

While the shareholders may exert net control over decision-making within the firm, the bank must continue to retain significant influence over such decisions if the close firm-bank relationship is
to successfully reduce information problems. A testable implication of the standard corporate
governance hypothesis, as a mechanism for reducing information problems between the firm and
the bank, is that the underlying corporate control regime should be responsive to exogenous
economic and financial shocks.

Our empirical results provide quantitative support for the Japanese financial system's ability to
reduce information problems between the firm and the main bank. Corporate control appears to
respond to changes in the economic and financial environment, both over time as well as across
firms, providing empirical evidence in favor of the corporate governance hypothesis, as the
mechanism through which such information problems are reduced. As mentioned earlier, the
close firm-bank relationship also does not appear to distort the investment-portfolio toward
excessively safe projects in the Japanese pharmaceutical industry.
References


Figure 1: Product-market / Financial-market Equilibrium

(a) Entrepreneur control

(b) Bank control
<table>
<thead>
<tr>
<th>Year</th>
<th>Share</th>
<th>P(AC-AB)</th>
<th>P(CV-AB)</th>
<th>OWN</th>
<th>TOT</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.253</td>
<td></td>
<td></td>
<td>5.100</td>
<td>8.229</td>
<td>1.322</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td></td>
<td></td>
<td>(0.748)</td>
<td>(2.799)</td>
<td>(0.413)</td>
</tr>
<tr>
<td>1981</td>
<td>0.232</td>
<td>0.370</td>
<td>0.391</td>
<td>5.467</td>
<td>9.102</td>
<td>2.433</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td></td>
<td></td>
<td>(1.219)</td>
<td>(2.673)</td>
<td>(3.040)</td>
</tr>
<tr>
<td>1982</td>
<td>0.212</td>
<td>0.378</td>
<td>0.398</td>
<td>5.400</td>
<td>10.587</td>
<td>1.941</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td></td>
<td></td>
<td>(1.124)</td>
<td>(3.155)</td>
<td>(2.070)</td>
</tr>
<tr>
<td>1983</td>
<td>0.164</td>
<td>0.474</td>
<td>0.476</td>
<td>5.150</td>
<td>11.892</td>
<td>1.241</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td></td>
<td></td>
<td>(0.887)</td>
<td>(3.655)</td>
<td>(0.651)</td>
</tr>
<tr>
<td>1984</td>
<td>0.151</td>
<td>0.611</td>
<td>0.600</td>
<td>5.100</td>
<td>11.741</td>
<td>1.311</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td></td>
<td></td>
<td>(0.768)</td>
<td>(4.442)</td>
<td>(0.674)</td>
</tr>
<tr>
<td>1985</td>
<td>0.143</td>
<td>0.761</td>
<td>0.812</td>
<td>5.086</td>
<td>13.011</td>
<td>1.123</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td></td>
<td></td>
<td>(0.669)</td>
<td>(4.984)</td>
<td>(0.506)</td>
</tr>
<tr>
<td>1986</td>
<td>0.166</td>
<td>0.775</td>
<td>0.854</td>
<td>4.675</td>
<td>12.833</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td></td>
<td></td>
<td>(0.871)</td>
<td>(6.149)</td>
<td>(0.508)</td>
</tr>
<tr>
<td>1987</td>
<td>0.148</td>
<td>0.567</td>
<td>0.588</td>
<td>4.500</td>
<td>14.160</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td></td>
<td></td>
<td>(0.693)</td>
<td>(7.017)</td>
<td>(0.403)</td>
</tr>
<tr>
<td>1988</td>
<td>0.133</td>
<td>0.653</td>
<td>0.649</td>
<td>4.438</td>
<td>15.435</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td></td>
<td></td>
<td>(0.663)</td>
<td>(7.324)</td>
<td>(0.442)</td>
</tr>
<tr>
<td>1989</td>
<td>0.131</td>
<td>0.667</td>
<td>0.687</td>
<td>4.475</td>
<td>17.317</td>
<td>1.057</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td></td>
<td></td>
<td>(0.650)</td>
<td>(8.103)</td>
<td>(0.431)</td>
</tr>
<tr>
<td>1990</td>
<td>0.142</td>
<td>0.679</td>
<td>0.653</td>
<td>4.525</td>
<td>18.965</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td></td>
<td></td>
<td>(0.523)</td>
<td>(8.737)</td>
<td>(0.415)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Firm</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.121</td>
<td></td>
<td></td>
<td>4.343</td>
<td>8.800</td>
<td>1.790</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td></td>
<td></td>
<td>(0.162)</td>
<td>(1.974)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>2</td>
<td>0.157</td>
<td></td>
<td></td>
<td>4.908</td>
<td>6.292</td>
<td>1.357</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td></td>
<td></td>
<td>(0.239)</td>
<td>(1.615)</td>
<td>(0.276)</td>
</tr>
<tr>
<td>3</td>
<td>0.333</td>
<td></td>
<td></td>
<td>4.317</td>
<td>16.974</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
<td>(0.159)</td>
<td>(4.668)</td>
<td>(0.084)</td>
</tr>
<tr>
<td>4</td>
<td>0.158</td>
<td></td>
<td></td>
<td>5.633</td>
<td>13.565</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td></td>
<td></td>
<td>(0.618)</td>
<td>(6.674)</td>
<td>(0.187)</td>
</tr>
<tr>
<td>5</td>
<td>0.138</td>
<td></td>
<td></td>
<td>4.942</td>
<td>17.663</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td></td>
<td></td>
<td>(0.051)</td>
<td>(7.067)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>6</td>
<td>0.073</td>
<td></td>
<td></td>
<td>5.940</td>
<td>13.170</td>
<td>2.711</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td></td>
<td></td>
<td>(1.042)</td>
<td>(3.820)</td>
<td>(2.555)</td>
</tr>
<tr>
<td>7</td>
<td>0.174</td>
<td></td>
<td></td>
<td>3.080</td>
<td>5.501</td>
<td>0.287</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td></td>
<td></td>
<td>(0.179)</td>
<td>(0.539)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>8</td>
<td>0.175</td>
<td></td>
<td></td>
<td>4.733</td>
<td>17.293</td>
<td>1.290</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td>(0.123)</td>
<td>(6.263)</td>
<td>(0.300)</td>
</tr>
</tbody>
</table>

**NOTE:**
Share = antibiotic-share
P(AC-AB)/P(CV-AB) = anticancer-antibiotic / cardiovascular-antibiotic sales per product OWN
= main bank ownership (%)
TOT = total R&D expenditure in (thousands of) million yen
R = debt-equity ratio
Standard deviations are in parentheses.
Table 2: Identifying corporate control

<table>
<thead>
<tr>
<th></th>
<th>P(AC-AB)</th>
<th>P(AC-AB)</th>
<th>P(CV-AB)</th>
<th>P(CV-AB)</th>
<th>Year dummies</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>-0.019</td>
<td>-0.022</td>
<td>-0.017</td>
<td>-0.022</td>
<td>-0.026</td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>OWN</td>
<td>0.032</td>
<td>0.033</td>
<td>0.030</td>
<td>0.032</td>
<td>0.043</td>
</tr>
<tr>
<td>(0.029)</td>
<td>(0.024)</td>
<td>(0.029)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>TOT</td>
<td>0.008</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>-0.106</td>
<td>-1.359</td>
<td>-0.067</td>
<td>-1.248</td>
<td></td>
</tr>
<tr>
<td>(0.065)</td>
<td>(0.455)</td>
<td>(0.052)</td>
<td>(0.440)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>P²</td>
<td>1.045</td>
<td>-0.901</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.356)</td>
<td>(0.318)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm1</td>
<td>0.081</td>
<td>0.083</td>
<td>0.077</td>
<td>0.082</td>
<td>0.099</td>
</tr>
<tr>
<td>(0.042)</td>
<td>(0.040)</td>
<td>(0.043)</td>
<td>(0.040)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>Firm2</td>
<td>0.054</td>
<td>0.052</td>
<td>0.054</td>
<td>0.051</td>
<td>0.063</td>
</tr>
<tr>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.035)</td>
<td>(0.033)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>Firm3</td>
<td>0.172</td>
<td>0.170</td>
<td>0.171</td>
<td>0.170</td>
<td>0.167</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Firm4</td>
<td>-0.031</td>
<td>-0.033</td>
<td>-0.029</td>
<td>-0.033</td>
<td>-0.030</td>
</tr>
<tr>
<td>(0.028)</td>
<td>(0.026)</td>
<td>(0.028)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Firm5</td>
<td>-0.065</td>
<td>-0.065</td>
<td>-0.064</td>
<td>-0.065</td>
<td>-0.063</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Firm6</td>
<td>-0.067</td>
<td>-0.063</td>
<td>-0.067</td>
<td>-0.064</td>
<td>-0.064</td>
</tr>
<tr>
<td>(0.034)</td>
<td>(0.030)</td>
<td>(0.034)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td>Firm7</td>
<td>0.185</td>
<td>0.179</td>
<td>0.182</td>
<td>0.177</td>
<td>0.202</td>
</tr>
<tr>
<td>(0.078)</td>
<td>(0.073)</td>
<td>(0.080)</td>
<td>(0.073)</td>
<td>(0.067)</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>-0.015</td>
<td>-1.208</td>
<td>-0.017</td>
<td>-2.315</td>
<td></td>
</tr>
<tr>
<td>(0.006)</td>
<td>(3.886)</td>
<td>(0.005)</td>
<td>(4.173)</td>
<td>(0.067)</td>
<td></td>
</tr>
<tr>
<td>Year²</td>
<td>0.000</td>
<td>-0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.042)</td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.043)</td>
</tr>
<tr>
<td>1982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.033)</td>
</tr>
<tr>
<td>1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.027)</td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.025)</td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.020)</td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.015)</td>
</tr>
<tr>
<td>1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.011)</td>
</tr>
<tr>
<td>CONST</td>
<td>30.715</td>
<td>1209.979</td>
<td>33.393</td>
<td>2309.045</td>
<td>0.226</td>
</tr>
<tr>
<td>(10.811)</td>
<td>(3857.7)</td>
<td>(10.460)</td>
<td>(4144.0)</td>
<td>(10.125)</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>0.157</td>
<td>0.164</td>
<td>0.171</td>
<td>0.167</td>
<td>0.033</td>
</tr>
<tr>
<td>R²</td>
<td>0.799</td>
<td>0.823</td>
<td>0.794</td>
<td>0.823</td>
<td>0.809</td>
</tr>
<tr>
<td>N</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>77</td>
</tr>
</tbody>
</table>

P(AC-AB)/P(CV-AB) = anticaner-antibiotic / cardiovascular-antibiotic sales per product
Newey-West standard errors are in parentheses in Table 2 and Table 3. Q: Box-Pearson Q-statistic.
### Table 3: Stability of the debt-equity and bank ownership coefficients

<table>
<thead>
<tr>
<th></th>
<th>Over time</th>
<th>Across firms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OWN</td>
<td>R</td>
</tr>
<tr>
<td>R</td>
<td>-0.049</td>
<td>-</td>
</tr>
<tr>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R*d1</td>
<td></td>
<td>-0.031</td>
</tr>
<tr>
<td>(0.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R*d2</td>
<td></td>
<td>-0.072</td>
</tr>
<tr>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R*d3</td>
<td></td>
<td>-0.024</td>
</tr>
<tr>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R*d0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN</td>
<td>-0.054</td>
<td>-</td>
</tr>
<tr>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN*d1</td>
<td>0.100</td>
<td>-</td>
</tr>
<tr>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN*d2</td>
<td>0.010</td>
<td>-</td>
</tr>
<tr>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN*d3</td>
<td>0.063</td>
<td>-</td>
</tr>
<tr>
<td>(0.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OWN*d4</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOT</td>
<td>0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>Firm1</strong></td>
<td>0.079</td>
<td>0.132</td>
</tr>
<tr>
<td>(0.024)</td>
<td>(0.039)</td>
<td>(0.030)</td>
</tr>
<tr>
<td><strong>Firm2</strong></td>
<td>0.031</td>
<td>0.074</td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.031)</td>
<td>(0.022)</td>
</tr>
<tr>
<td><strong>Firm3</strong></td>
<td>0.151</td>
<td>0.162</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.023)</td>
<td>(0.017)</td>
</tr>
<tr>
<td><strong>Firm4</strong></td>
<td>-0.063</td>
<td>-0.038</td>
</tr>
<tr>
<td>(0.015)</td>
<td>(0.025)</td>
<td>(0.017)</td>
</tr>
<tr>
<td><strong>Firm5</strong></td>
<td>-0.075</td>
<td>-0.073</td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.016)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Firm6</strong></td>
<td>-0.059</td>
<td>-0.049</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.026)</td>
<td>(0.015)</td>
</tr>
<tr>
<td><strong>Firm7</strong></td>
<td>0.154</td>
<td>0.232</td>
</tr>
<tr>
<td>(0.063)</td>
<td>(0.085)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>1980</td>
<td>-0.020</td>
<td>0.225</td>
</tr>
<tr>
<td>(0.144)</td>
<td>(0.085)</td>
<td>(0.142)</td>
</tr>
<tr>
<td>1981</td>
<td>-0.017</td>
<td>0.223</td>
</tr>
<tr>
<td>(0.142)</td>
<td>(0.084)</td>
<td>(0.140)</td>
</tr>
<tr>
<td>1982</td>
<td>-0.063</td>
<td>0.178</td>
</tr>
<tr>
<td>(0.136)</td>
<td>(0.076)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>1983</td>
<td>0.335</td>
<td>0.161</td>
</tr>
<tr>
<td>(0.116)</td>
<td>(0.070)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>1984</td>
<td>0.315</td>
<td>0.140</td>
</tr>
<tr>
<td>(0.114)</td>
<td>(0.068)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>1985</td>
<td>0.291</td>
<td>0.108</td>
</tr>
<tr>
<td>(0.112)</td>
<td>(0.065)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>1986</td>
<td>0.293</td>
<td>0.117</td>
</tr>
<tr>
<td>(0.114)</td>
<td>(0.067)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>1987</td>
<td>0.036</td>
<td>0.052</td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.023)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>1988</td>
<td>0.016</td>
<td>0.027</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>1989</td>
<td>0.004</td>
<td>0.007</td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.009)</td>
</tr>
<tr>
<td><strong>CONST</strong></td>
<td>-0.231</td>
<td>-0.305</td>
</tr>
<tr>
<td>(0.144)</td>
<td>(0.138)</td>
<td>(0.147)</td>
</tr>
<tr>
<td>Q</td>
<td>0.110</td>
<td>0.020</td>
</tr>
<tr>
<td>(0.894)</td>
<td>(0.822)</td>
<td>(0.891)</td>
</tr>
</tbody>
</table>

N: 77

Note: 1 = 1 if YEAR = 1980-82, 2 = 1 if YEAR = 1983-86, 3 = 1 if YEAR = 1987-90, 4 = 1 if R > AVG(R), 0 = 1 if firm = firm3, firm4, firm5, firm7
Chapter 2:  
Milk Supply Behavior in India:  
Data Integration, Estimation and Implications for Dairy Development

1. Introduction
The performance of the Indian dairy sector over the past two decades has been extremely impressive. Milk production grew at an average annual rate of 4.2% from 1967 to 1986 and India is today the world’s third largest producer of milk. A vast dairy infrastructure system was established in this period with the assistance of the Indian government and foreign donor agencies. It consists essentially of a network of privately owned and democratically organized milk producers’ cooperative societies. The initial objective of the cooperative societies was to collect and market milk, providing remunerative prices to the producers. Over the years their role has been extended to include dairy development, specifically the improvement in the quality of the milch herd. The dairy development programs have included veterinary and animal husbandry extension services, artificial insemination (AI) schemes, and livestock-disease vaccination programs.

This combination of rapid production growth and a unique industrial organization created substantial interest in the Indian dairy industry among development agencies and academics. The response in the literature to the dairy development effort however, has not been altogether positive (see Alderman et. al., 1987, for a review). Most of the dairy development programs, viewed in isolation, appear to have failed to satisfy their stated objectives. For instance, the Artificial Insemination (AI) program, which was the primary instrument in the National Dairy Development Board’s (NDDB’s) effort to create a National Milch Herd of genetically improved animals, has been recognized to have been technically flawed (Shah, 1987). While an efficient

---

12 This chapter is co-authored with Kirit S. Parikh.

13 The National Dairy Development Board (NDDB) is the apex body in the cooperative structure and is responsible for the coordination of the development programs associated with the cooperative system. Operation Flood I (1970-81) attempted to organize a milk collection and marketing infrastructure for 18 selected districts of the country. Operation Flood II (1979-86) covered a greater area (40 districts) and concentrated on increasing the genetic potential of milch animals through cross-breeding in addition to improving the dairy infrastructure. The scheme to create a National Milch Herd was one of the stated objectives of Operation Flood II.
AI delivery system has been established over an extensive area, the necessary support programs have not proceeded apace. Although it is recognized that the adoption of cross-bred cows is severely hindered by a general lack of access to green fodder, the aggregate supply of cultivated fodder has also not been increased substantially over the past three decades. In addition, the susceptibility of cross-bred cows to various diseases, especially Foot-and-Mouth Disease (FMD) which is endemic in India, continues to remain a problem. An ambitious vaccination program to eradicate FMD was launched by the NDDB in the 1980s and eventually covered selected districts in the states of Karnataka, Kerala, and Tamil Nadu. The FMD program was extremely successful initially and the incidence of the disease declined dramatically in many of the target-districts. This success was short-lived, however. The vaccination coverage later declined in many areas, leading to a collapse in herd-immunity and a re-emergence of the disease.

In the absence of any other well defined determinant, the growth in milk production has been typically attributed to increased feeding of the milch herd. The drawbacks of feed-intensive growth are well known. Such a strategy cannot be sustained in the long-run, given the limited production potential of the existing milch stock and constraints on feed supply. It is also associated with adverse distributional impacts. In this regard, it has been argued that the growth in milk production has resulted in the diversion of scarce agricultural products from human consumption to the dairy sector, adversely affecting poor households. In addition, there is a perception that the benefits of this growth have been concentrated among richer buffalo and cross-bred cow owners, excluding indigenous cow owners from the development process.

While much of the literature has assumed that increased feeding is the chief determinant of the growth in milk production, there is little evidence to support this position. While many of the dairy development programs do appear to have failed, it is entirely possible that intangible externalities associated with these programs provided information and increased awareness among livestock owners. In addition, the last three decades witnessed significant rural development, especially in agriculture, which may have had spill-overs in the dairy sector as well.

The principal objective of this paper is to decompose the sources of the growth in milk production. Our approach is very simple. We develop and estimate an aggregate milk supply model for India, using the available data in a consistent framework. In our model, milk
production is a function of the number of milch animals, the quantities of various feeds provided to the mature animal, the "health" status of the animal determined by the quantities fed to the animal when it was young, and the quality of the animal. The shares of the growth in milk production due to feed and technological progress can then be computed from the estimated model.

The bovine diet consists of concentrate feeds, such as bran, coarse grains, oilseeds, and pulses, in addition to green and dry fodder. Concentrate feeds are actively traded in contrast with fodder, for which markets are not well developed (Shah et al., 1980). In developing our model we assume that these feeds are allocated to both mature and young animals in a manner that is consistent with the livestock owner's profit-maximizing behavior. Milk supply thus responds to the relative prices of milk and concentrate feeds. Fodder is provided to the animal on the basis of its availability. Milk supply therefore also responds to changes in the supply of fodder per animal and is consistent with aggregate fodder supply in each year.

Improvement in the quality of the animal reflects technological progress in milk production. Technological progress occurs either through improvement in the quality of the existing stock or due to cross-breeding, which increases the genetic potential of the herd. As we discussed earlier, technological progress may result from focussed interventions such as veterinary and animal husbandry extension services, AI schemes, and livestock-disease vaccination programs. It could also be a response to more diffuse factors, such as general rural development for instance. The decomposition of technological progress in milk production into its various components is beyond the scope of this paper. In the absence of any other suitable determinant we include the number of cooperative societies as the sole measure of technological progress in our model.14 This variable is associated with most of the directed interventions in the dairy sector. As we will discuss below, it may also be associated with more diffuse development factors since the cooperative network provides a natural channel for the dissemination of information as well as an infrastructure base for the adoption of new technology.

14 There have been two major interventions over the past decades in the Indian dairy industry - by the government and within the cooperative system. The government effort, which included the Intensive Cattle Development Program (ICDP), is generally considered to have had little impact on dairy development (Shah, 1987).
Our econometric procedure is potentially of general relevance to empirical work with LDC data. The model is estimated using aggregate time-series data from 1961 to 1986, separately for cows and buffaloes. Multicollinearity is a problem often associated with time-series data. This problem is particularly severe with aggregate LDC data since sample sizes are typically small. While multicollinearity does not lead to inconsistent estimates in principle, in practice it can give rise to wildly fluctuating estimates in small samples. The use of extraneous information to reduce the number of parameters to be estimated is often suggested as a solution to the multicollinearity problem. The drawback of this approach is that it could conceivably lead to specification bias. In general it is not clear which of these two approaches is to be preferred.

In our paper we estimate both an unrestricted as well as a restricted model. The unrestricted model provides consistent estimates in principle but gives rise to unrealistic point estimates in our case, possibly because of multicollinearity. The estimated coefficients of the restricted model on the other hand, are consistent with economic theory and the results from various cross-sectional micro-studies conducted over the estimation period. Since we derive a structural model it is also possible to test the over-identifying restrictions. In our particular case we are forced to reject the null hypothesis for some of these restrictions, implying specification bias in the restricted estimates. Thus, for our model neither the restricted nor the unrestricted specification unambiguously dominates the other in general. We do however, present additional independent evidence that supports the restricted model. Nevertheless, we report results for both models and show that our main qualitative conclusions are unaffected by the choice of econometric specification.

Technological progress and the increased use of feed are specified as the sources of growth in milk production in our model. Technological progress is found to be responsible for a large share of this growth in recent years. This share has also increased over time. This suggests that the preoccupation with feed utilization and feed supply accounting in the literature may have been somewhat misguided. What seems of greater interest are the non-feed determinants of the growth in milk production.
The large share of this growth attributed to the cooperative system in our model is somewhat surprising. Villages with cooperatives account for less than 15% of total milk production. However, as we discussed earlier, the cooperative system may provide a natural channel for the dissemination of information as well as an infrastructure base for the adoption of new technology. It may also facilitate the social learning process that is often associated with technology diffusion. In that sense, its scope extends beyond cooperative villages, and milk collection data or the results of individual development schemes underestimate its importance in dairy development. While additional research is clearly necessary to support this hypothesis, it has potentially important implications for dairy development strategy.

The paper is organized in six sections. Section 2 describes the model and section 3 discusses data sources and construction. Section 4 gives details of the estimation procedure and validates the model. Section 5 describes the results of the model and section 6 concludes the paper, summarizing the main results and suggesting areas for future research.

2. The Model

Cows and buffaloes together account for most of India’s milk production. Cows may be classified as indigenous and cross-bred. Indigenous cows are randomly bred, have very low milk yields, and can produce milk on a poor diet. There are roughly 55 million breeding indigenous cows and this number has not changed significantly over the past three decades. Genetically improved cows are produced by cross-breeding indigenous cows with improved breeds (typically Holstein-Friesian and Jersey). While these animals are more efficient at converting feed to milk than indigenous cows and buffaloes, they are not as hardy and require large quantities of expensive concentrate feeds and cultivated fodder to provide milk at their full potential. The impact of these animals on national milk production has been only a fraction of what was envisaged by policy-

---

15 Based on milk procurement data, the impact of the cooperative system would appear to be quite limited. For instance, only 6.5% of total milk production was procured under Operation Flood in 1986. Since 72% of the cooperatives were covered by Operation Flood by this point, this would imply that roughly 9% of milk production was collected from the entire cooperative system in 1986. Including self-consumption by cooperative households, the share of the cooperative system in milk production would probably not exceed 15%.
makers. Buffaloes on the other hand, are well suited to the Indian environment, producing more milk than indigenous cows but requiring more nutritious food. Their numbers have been increasing steadily over the past three decades and there are now roughly 30 million breeding animals.

We begin with a simple identity. Total national milk production is the product of herd size, the probability that animals are lactating (proportion of breeding animals in-milk) and the yield of lactating animals (Alderman, 1987 and Nair, 1985). The model is specified, and later estimated, separately for cows and buffaloes.

\[ M = N \cdot \frac{L}{N} \cdot \frac{M}{L} \]  

(1)

where

\( M \) = milk production (cows or buffaloes)  
\( N \) = number of breeding animals  
\( L \) = number of lactating animals.

We consider constant returns to scale in milk production with respect to the number of animals in the herd to be a reasonable assumption. The yield-input relationship does not depend on the number of animals and hence it is not possible to determine both the technique (level of input) and the scale of production (number of animals). The number of animals \( (N) \) is thus taken to be exogenous and the average yield \( (L/N \cdot M/L) \) is the dependent variable in our model. The average yield for the herd is obtained as,

\[ YLD_t = \sum_{i = 1}^{15} YLD_i \]  

(2)

where

\( YLD \) = average milk yield for cow or buffalo (Kg./day)  
\( t \) = index for current year

---

\(^{16}\) Operation Flood II (1979-86) called for the creation of a "National Milch Herd" of approximately 20-25 million cross-bred cows and upgraded buffaloes (Halse, 1979). However, the 1982 Livestock Census (GOI, 1982) recorded only 2.7 million cross-bred cows (4.6% of the total number of breeding cows). Given the general impression that OF has not contributed appreciably to the improvement in the genetic potential of buffaloes either, this implies that the OF cross-breeding program at least, has fallen far short of its expectations (Nair, 1985).
\[ \tau = \text{index for age (4-15 years)}^{17} \]
\[ S_t^\tau = \text{proportion of mature animals aged } \tau \text{ in year } t. \]

Feed quantities are assumed to be the only variable inputs determining average yield. Other variables such as labor and management, have been suggested in the literature. However, simple increases in labor intensity or herd management effort are not generally believed to be particularly effective in increasing milk yield, and feed is thus typically considered to be the principal scarce input in milk production (Singh and Jha, 1975). Improvements in animal husbandry practices and management techniques are important and effective however, and are included under technological progress in our model.

It has been recognized that milk yield depends on feeding levels in the current lactation as well as in the past (Dean, 1960). We assume that feed quantities provided to the animal when it was young influence its production potential. Milk production is thus a function of feed given in the current year as well as the feed the animal received in its formative years.

A Cobb-Douglas specification is chosen for the yield function. This specification is characterized by the restrictive properties of constant elasticity of scale and unitary elasticity of substitution. However, it is used to estimate milk yield in most micro-level studies in India.\textsuperscript{18} As described later, we will need to exogenously specify some of the elasticities obtained in these studies while estimating our model. We therefore use the same form to maintain consistency in specification.

Based on the above discussion the average yield for animals who are \( \tau \) years old in year \( t \) is specified as,

\[
YLD_t^\tau = A \cdot (Qy_{t-\tau+1} \cdot Qy_{t-\tau+2} \cdot Qy_{t-\tau+3})^\gamma \cdot (Qm_t)^\delta \cdot e^{\lambda TECH},
\]

where

\[ Qy_{t-\tau+1} = \text{quantity fed to the animal when it was one year old.} \]

\textsuperscript{17} We assume that the animal remains unproductive (young) for three years (as in the Livestock Census) and has a service life of twelve years (NDDB, 1980).

Similarly for \( Q_{y_{t+2}} \) and \( Q_{y_{t+3}} \). Notice that all young animals are assumed to be fed the same quantity in a given year \( t \) regardless of their age (1-3 years).

\[ Q_m = \text{quantity fed to mature animal} \]

TECH = technological progress variable (number of cooperative societies in our model)

\( \alpha, \beta = \text{production elasticities} \)

\( \lambda = \text{technological progress coefficient} \)

\( A = \text{constant term, to account for the change in units} \).

The feed quantities in equation (3) are measured in "equivalent nutrition" units and consist of a combination of various feed-types. As described earlier, feeds can be broadly divided into concentrates (bran, coarse grains, oilseeds, pulses) and fodder (green and dry). We stipulate a Cobb-Douglas function to "produce" the equivalent nutrition from the different feed-types. Hence,

\[ Q_m = (D_m)\delta \cdot (G_m)\gamma \cdot \prod_i (C_m_i)^{\epsilon} \tag{4} \]

\[ Q_y = (D_y)\delta \cdot (G_y)\gamma \cdot \prod_i (C_y_i)^{\epsilon} \tag{5} \]

where

\( D, G, C = \text{dry fodder, green fodder, concentrate quantities respectively (Kg./day)} \)

\( i = \text{concentrate-type index (bran, coarse grains, oilseeds, pulses)} \)

\( \delta, \gamma, \epsilon = \text{production elasticities for dry fodder, green fodder, concentrates respectively} \)

\( m, y \) refer to mature and young animal respectively

\( t, \tau \) are ignored to reduce notational clutter.

Note that the production elasticities in equations (4) and (5) are the same for both young and mature animals. This assumption implies that nutrients from a given feed are equally accessible nutritionally to both young and mature animals, leaving us with greater degrees of freedom in the estimation procedure.

The model is difficult to estimate since we do not have data on the numbers, quality, and feed consumption, for mature animals of different ages. We therefore simplify the yield function by assuming,

**A1.** All mature animals are fed the same amount in a given year, \( Q_{m_t}' = Q_{m_t} \).
The nutrition quantity fed to young animals has been increasing with a geometric trend over time, \((Q_{Y_{t+1}}, Q_{Y_{t+2}}, Q_{Y_{t+3}}) = (Q_{Y_{t+3}})^3\).

These assumptions lead to a simplified average yield function as follows,

\[
YLD_t = A \left( \sum_{t=2}^{12} S_t^{T-2} \cdot (Q_{Y_{t-r}})^{3a} \right) \cdot (Qm)^\varrho \cdot e^{\lambda TECH}, \tag{6a}
\]

where

\(T = \) lag index in the model.

We make one additional simplifying assumption,

**A3.** \(\Sigma_t S_t^{T+2} \cdot (Q_{Y_{T-r}})^{3a} = (Q_{Y_{T-r}})^{3a}\)

where \(r\) is an unspecified lag period, \(r \in [2,13]\).

The last two simplifying assumptions are made for convenience: \((Q_{Y_{t+1}}, Q_{Y_{t+2}}, Q_{Y_{t+3}})\) is replaced by \((Q_{Y_{t+3}})^3\) and \(S_t^{T+2} \cdot (Q_{Y_{T-r}})^{3a}\) is effectively replaced by \((1/12)(Q_{Y_{T-r}})^{3a}, \forall T \in [2,13]\). This leads to specification bias unless the relationships are exactly satisfied. The restrictions implied by these assumptions are testable however, and as we shall see the advantages of these simplifications are significant and may exceed the cost of the possible mis-specification.

Using the last simplifying assumption to modify equation (6a), substituting equations (4) and (5) in (6a), and letting \(a = 3\alpha/\beta, d = \beta \delta, g \equiv \beta \gamma, e_i \equiv \beta \varepsilon_i\), we obtain the final yield specification as,

\[
YLD_t = A \cdot (D_{Y_{t-r}})^{3d} \cdot (Q_{Y_{t-r}})^{3g} \cdot [(C_{Y_{t-r}})^{a} \cdot (Dm)^{d} \cdot (Gm)^{g} \cdot [(Cm)^{e_i} \cdot e^{\lambda TECH}, \tag{6b}
\]

We require input quantities in each year to estimate the parameters of the yield function. As discussed earlier, fodder is not actively traded in contrast with concentrate feeds, so prices cannot be used to derive the demand for fodder. We assume that dry fodder is allocated to satisfy the maintenance function while concentrate feeds are "added-on" to maximize profits from milk.
production.\textsuperscript{19} The maintenance requirement (derived from Sen and Ray, 1971) summed over all bovines (cows, buffaloes, bullocks, buffalo males, and young animals) invariably exceeds the total dry fodder supply. Since the maintenance of all productive animals is equally important we assume that livestock owners allocate dry fodder as the same proportion of the respective maintenance requirement for each bovine-type such that the total fodder supply is completely utilized in each year. We cannot assume that green (cultivated) fodder is provided to satisfy the maintenance requirement since it is a relatively valuable feed. The allocation scheme is therefore based on the relative consumption of cross-bred cows, buffaloes and indigenous cows obtained from the NCAER 1988-89 Survey (NCAER, 1990). Since in India fodder is consumed by bovines only (Amble et al., 1965), the pre-allocation rules allow us to derive the fodder quantity allocated to each mature and young milch animal (Dm,Gm,Dy,Gy), separately for cows and buffaloes, subject to the fodder availability constraint in each year. We cannot derive concentrate quantities per animal from total supply in a similar manner since they are consumed by humans, poultry, and hogs in addition to bovines (Ibid.) However, since concentrates are actively traded, it may be reasonable to assume that the first-order profit-maximizing conditions describe their use in milk production. Concentrate quantities fed to both the mature animal (Cm,) and the young animal (Cy), can be derived from the first-order conditions as milk and input prices are known.\textsuperscript{20}

To maximize profit each concentrate feed is fed to the animal until its marginal value product is equal to its price. For the mature animal the first-order conditions are obtained from equation (6b) as,

\textsuperscript{19} Apart from the production of milk, feeds are required for the maintenance of the animal as well. Vaidyanathan (1988, p.77) distinguishes between these two functions.

\textsuperscript{20} We construct separate series for cow and buffalo milk prices in our paper. The price of milk in India reflects the value of its fat content (Mergos and Slade, 1987) and hence, the prices of cow milk and buffalo milk may be derived separately from the general price of milk since their fat contents are known. The assumption of fat-based pricing is not strictly accurate however, since some weightage has been given to non-fat solids (SNF) since the 1970s in many districts (Shah et al., 1990). This weightage has been ignored in our model, since it varies quite widely, slightly under-valuing cow milk prices.
\( (P)_i = \frac{e_i \cdot \text{YLD}_i \cdot PM_i}{(Cm)_i} \)  

(7)

where

\( P \) = price of concentrate (Rs./Kg.)

\( PM \) = price of milk (Rs./Kg.).

Note that we have not considered the value of any calves produced in the future by the animal in the first order condition. The primary function of indigenous cows has traditionally been to produce bullocks (Vaidyanathan, 1988) and at least for those animals, the value of their offspring would be an important component of the profit function. By omitting this component we make the implicit assumption that the value of the calves is independent of the quantity of variable inputs provided to the mature animal.

The first-order conditions for the young animal are much more complex as the price of the concentrate feed is equated to the present value of the future stream of marginal value products due to the young animal feed. Hence, substituting from equation (5) in (6a), letting \( a = 3\alpha/\beta \), \( e_i = \beta e_i \), and assuming perfect foresight,

\[(P)_i = \frac{a \cdot e_i}{(Qy)_i} \cdot A \cdot (Qy)_i^{3\alpha} \sum_{T=2}^{T+2} \frac{S_i \cdot (Qm \cdot t)^{\beta} \cdot e^{\lambda TECH \cdot t} \cdot PM_i \cdot t}{I^T} \]

(8a)

where, \( I \) = time discount rate.\(^{21}\)

The assumption \( \Sigma_T S_i \cdot (Qy \cdot t)^{3\alpha} \cdot (Qy \cdot t)^{3\alpha} = (Qy \cdot t)^{3\alpha} \), that we made earlier, however, allows us to simplify equation (8a) considerably.\(^{22}\) From equation (6b), and assuming perfect foresight once more,

As we shall describe later in section 4 we were unable to estimate individual fodder elasticities, possibly due to multicollinearity. These elasticities are therefore exogenously specified, allowing

\(^{21}\) Crotty (1980, p.182) estimates this as roughly 1.4 for Indian agriculturalists. We do not require this value for the estimation procedure since, \( I \), is absorbed in a constant term which is estimated in its reduced form.

\(^{22}\) The simplified yield specification modifies the livestock owner’s young feed allocation decision as well. He now predicts the future yield and milk price for a single representative year only. This is in contrast with the original specification with which he must predict variable levels for 12 future years in equation (8a).
\begin{equation}
(P)_{r} = \frac{a.e_{i}}{(Cy)_{r}} \cdot (Qy_{t,r})^{\alpha} \cdot (Qm_{r,r})^{\beta} \cdot e^{\text{TECH}_{r,r}} \cdot PM_{r,r} = \frac{a.e_{i}}{(Cy)_{r}} \cdot \frac{YLD_{r,r} \cdot PM_{r,r}}{I'}
\end{equation}

us to construct composite young and mature fodder quantities,
\[ Fy_{t,r} = (Dy_{t,r})^{a} \cdot (Gy_{t,r})^{b}, \quad Fm_{r} = (Dm_{r})^{c} \cdot (Gm_{r})^{d}. \]
Substituting for \((Cy)_{r}\) and \((Cy)_{t,r}\) from equations (7) and (8b), taking logs and collecting terms, we obtain the reduced form of the yield function from equation (6b) as,
\begin{equation}
ln(YLD_{r}) = \pi_{1} + \pi_{2} \ln Fy_{t,r} + \pi_{3} \ln Fm_{r} + \pi_{4} \ln PM_{r} - \sum_{i} (\pi_{5,i} + \pi_{6,i} \ln (P)_{i,r} + \pi_{7} \ln (P)_{t}) + \pi_{7} \text{TECH}_{r}
\end{equation}
where
\[ \pi_{1} = \frac{1}{1-(1+a) \sum_{i} e_{i}} \ln \left( \frac{a}{I'} \left[ \frac{\pi \sum_{i} e_{i} \left( 1^{1-a} \right)}{1-a \sum_{i} e_{i}} \right] \right) \]
\[ \pi_{2} = \frac{a}{1-(1+a) \sum_{i} e_{i}} \pi_{3} = \frac{1}{1-(1+a) \sum_{i} e_{i}} \]
\[ \pi_{4} = \frac{(1+a) \sum_{i} e_{i}}{1-(1+a) \sum_{i} e_{i}} \pi_{5,i} = \frac{ae_{i}}{1-(1+a) \sum_{i} e_{i}} \pi_{6,i} = \frac{e_{i}}{1-(1+a) \sum_{i} e_{i}} \pi_{7} = \frac{\lambda}{1-(1+a) \sum_{i} e_{i}} \]
Equation (6c) with an additive error term, is estimated in its reduced form using OLS, as well as in its structural form using Non-Linear Least Squares (NLLS).\(^{23}\) The details of the estimation procedure are provided in Section 4.

The principal advantage of the simplifying assumptions, \((Qy_{t+1}, Qy_{t+r}, 2, Qy_{t+r+2}, (Qy_{t+r+2})^{3}\) and \(\Sigma_{T_{r}} S_{r}^{t+2}(Qy_{t,T})^{3}= (Qy_{t,T})^{3}\), are now apparent. They allow us to express \((Cy)_{r}\) and \((Cm)_{r}\) as functions of the yield in year \(t\), \((YLD)\). We are thus able to derive \(YLD\), explicitly as a function of exogenous variables. In an earlier version of the paper (Munshi and Parikh, 1990), we estimated the model without this simplification using equation (8a). Consequently, we could not estimate the model in one-step and were forced to employ a recursive procedure to simultaneously

\(^{23}\) The presence of both quantities and prices on the right-hand-side of the yield function causes no problems. These variables are determined independently of the livestock owner’s behavior and are therefore, exogenous to the model. Our specification thus avoids the simultaneous equation bias that is typically associated with the production function approach (Lau and Yotopoulos, 1972).
estimate the yield function parameters and generate the concentrate quantities.24 It is difficult to interpret parameters estimated with this recursive procedure and hypothesis testing is also not entirely straightforward. We face no such difficulties with the model in its current simplified form.

3. Data Sources and Construction

The data required to estimate the model are not available at one place and have to be put together from a number of sources. The principal data sources are the India Database (Chandhok et. al., 1990) and the FAO Database (FAO, 1990). Both these databases are compiled from official Government of India (GOI) data. Other sources include the Report of the National Commission on Agriculture (GOI, 1976), the NCAER 1988-89 Livestock Survey (NCAER, 1990), National Livestock Census Reports from 1961 to 1982 (GOI, 1961, 1966, 1972, 1977, 1982), NDDB (1980), the Sen and Ray Feeding Standards (Sen and Ray, 1971), and various articles in the literature.

Time series for the following variables are required to estimate the model: number of animals (separately for indigenous and cross-bred cows, buffaloes, bullocks, buffalo males, cattle and buffalo young), milk production, fodder quantities, milk and concentrate feed prices, and the number of cooperative societies. In addition, the proportion of lactating indigenous cows in each year was required for subsequent analysis. Much of the required information cannot be directly obtained from the available data sources forcing us to make a number of simplifications and assumptions. Time trends were fitted to complete the series for some of the variables and spot values were assumed to hold for the entire time period under consideration in other cases. However, we have attempted to maintain consistency with all the available information in each case and a significant amount of time and effort was spent on the data construction (details of the

---

24 In the first iteration we assumed a set of parameter values and generated (Cm), and (Cy), from equations (7) and (8a) since YLD, is observed. We then estimated the parameters of the yield function from equation (6a), substituting expressions for Qy and Qm from equations (4) and (5) respectively. These parameter values were used to generate (Cm), and (Cy), in the second iteration. The iterative procedure was continued until the set of parameters used to generate the concentrate quantities was identical to the set of parameters estimated from the yield function in a given iteration. The procedure thus converged to a solution in which production and profit functions were consistent.
data construction as well as the data set used for the estimation of the model may be obtained from the authors on request).

4. Estimation and Validation of the Model

As we discussed earlier, we estimate two specifications of our model - with and without over-identifying restrictions. Equation (6c) can be estimated using OLS without placing any restrictions on the reduced form parameters. Alternatively, we can use Non-Linear Least Squares (NLLS) to estimate the structural coefficients (this is equivalent to OLS with non-linear restrictions). The restricted regression provides efficient but potentially inconsistent estimates. We consequently validate the structural model by comparing our estimates and the generated concentrate quantities with a priori theoretical predictions as well as the results from various cross-sectional micro-studies conducted over the estimation period. In addition, we conduct formal hypothesis tests of the over-identifying restrictions.

The validation and hypothesis testing is confined to the restricted model. The reduced form estimates have no economic interpretation and also cannot be used to generate concentrate quantities for comparison with the micro-studies. The unrestricted model is essentially used in this section as a bench-mark for the testing of the restricted model. While we do not present the reduced form estimates, some of the coefficients are used to derive the share of the growth in milk production due to technological progress in the next section.

We begin with a discussion on the estimation of the restricted model and its validation. As described above, the model is estimated using NLLS. Since we are using time-series data, we would expect autocorrelation in the error terms. As we shall see below in Table 1, the value of the Durbin-Watson statistic suggests that this is indeed the case in some instances. Consequently, we use the Newey-West robust estimator to obtain appropriate standard errors for hypothesis testing in the presence of autocorrelated disturbances with an unspecified structure.25

25 The only difficulty with applying the Newey-West estimator is that we must determine a lag L, a priori, beyond which autocorrelations are small enough to ignore. We specify L=4, after estimating the model separately with L∈[1,4] and determining that the standard errors are quite insensitive to the value of L.
We first estimated cow and buffalo yield using the specification in equation (6c). We were unable to obtain sensible estimated for all the parameters of the buffalo milk yield function. Some feed elasticities were negative and the value of the young-feed parameter, a, was significantly greater than unity, suggesting omitted variable bias. Consequently, we introduced two additional variables in the yield function for both cow and buffalo - a dummy variable to capture the "pre-cooperative" years, 1961-67, and a quadratic technological progress term (number of cooperative societies, squared) to capture the changing marginal effectiveness of the cooperative system as it expanded to cover almost the entire country.\(^{26}\)

As described earlier, we were forced to place additional restrictions on the feed elasticities in order to obtain estimates that were economically sensible. The fodder series are crudely derived from aggregate availability in each year in our model. The allocation procedure is based on the animals' "scientific" nutritional requirements without responding to variations in milk and concentrate feed prices. We were unable to obtain sensible estimates of any of the fodder elasticities in our model and so exogenously specified their values as the average of the estimates from a number of micro-level field studies, allowing us to construct the composite fodder quantities in equation (6c), \(F_y\) and \(F_m\).\(^{27}\) There is also significant multicollinearity in

\(^{26}\) The early 1960s were characterized by the dominance of the traditional milk-trader in milk collection and a sequence of famines (1964-65, 1965-66). The period 1961-67 saw negative average annual growth and sharp fluctuations in milk production for both cows and buffaloes, and is therefore considered as a distinct milk production regime in our model.

The cooperative system grew over the last three decades from roughly 3,000 societies in 1961 to over 60,000 in 1986. The system traces its origins to Kheda district of Gujarat state, a region with a long tradition of dairying and entrepreneurship. Over the years, as the cooperative system expanded into areas with less favorable natural endowments, we would expect the "marginal" cooperative society to have been less effective in increasing average milk yield. We consequently include the quadratic term to capture this changing marginal effect.

\(^{27}\) These studies include Lalwani (1989), Rao (1975), Ratnam (1982), Sankhayan and Joshi (1975). The elasticities for dry fodder and green fodder are found to be 0.0769 and 0.2396 respectively for cows. The corresponding values for buffaloes are obtained as 0.1457 and 0.3363.

We cannot compute a similar value for the concentrate elasticities since the micro-studies estimate a composite concentrate elasticity. It is not possible to derive a relationship between the individual concentrate elasticities that we require and the composite concentrate elasticity that is available.

Fodder quantities fed to the young animal and concentrate prices prior to 1961 are required since these variables are lagged by r years in the yield function. Unfortunately we do not have access to such data. Since levels for those variables stayed roughly constant from 1961 to 1966, we take the 1961 value to apply to all pre-1961 years. We check the sensitivity of the parameter estimates to the composite fodder quantities in equation (6c). The estimates are robust to the level of the composite fodder quantities, suggesting that the results of the model are not sensitive to the use of exogenously specified fodder elasticities and the extension for lagged variables that we
concentrate feed prices preventing us from estimating sensible individual elasticities. We restrict the model further by appropriately equating feed elasticities. The most reasonable combination appears to be bran/coarse grains and oilseeds/pulses \( e_b = e_c, e_o = e_p \).\(^{28}\) We present estimates with both restricted as well as unrestricted elasticities in Table 1 below. The fully restricted model is the only specification which provides sensible estimates for all feed parameters (positive concentrate elasticities and young-feed parameter, \( a \), less than unity).

Equation (6c) is specified with a lag period "\( r \)" to account for the feed that the animal received in its formative years. Since there is no particular lag that is preferred \textit{a priori}, we estimated the model separately for a number of lag years, \( r = 4, 5, 6, 7 \). Only \( r = 5 \) and \( r = 6 \) gave sensible estimates for all the parameters. The results of the estimation with \( r = 5 \) are presented in Table 1 below. While we do not report the estimates for \( r = 6 \), they are qualitatively similar to the coefficient estimates presented here.

All the estimated parameters in Table 1 (with \( e_b = e_c, e_o = e_p \)) have sensible signs and magnitudes. \( \lambda_2 \) is negative, implying that the marginal cooperative society is less effective in increasing milk yield. This is what we would expect as the cooperative system expands to cover areas with a lower potential for dairy development. \( \theta \) is greater for buffaloes than for cows, which is also expected, since the traditional milk-traders who dominated the pre-cooperative regime (1961-67) discriminated heavily in favor of buffalo milk due to its high fat content. The \( \lambda_1, \lambda_2, \theta \) estimates are highly statistically significant, the only exception being the \( \theta \) estimate for cows. This result supports our choice of technological progress variable (number of cooperatives) and the inclusion of the additional variables in the yield function (number of cooperatives, squared and the pre-cooperative regime dummy).

Given the tendency of variables to trend together over time we check that the importance of the cooperatives in our model is not due to spurious correlation arising from the omission of some unspecified explanatory variable. While the cooperative system appears to be the only likely non-

\(^{28}\) Inspection of the input price series shows oilseeds and pulses increasing rapidly over time, while bran and coarse grain remained relatively stable. Oilseeds and pulses also have higher protein and fat contents than the other two feed-types.
feed determinant of the growth in milk production from within the dairy sector, diffuse factors associated with general rural development could have played an important role as well. Since it is difficult to explicitly identify such factors, we include a time-trend as an additional explanatory variable to capture the effect of any unspecified determinants that may have increased milk yield over time. The results of this exercise are presented in Table 1. The time-trend coefficient is small in magnitude and statistically insignificant suggesting that no important variables have been omitted. The $\lambda_1$ and $\lambda_2$ estimates do not vary appreciably over the three specifications that we report. While the buffalo estimates are very stable, the cow estimates do vary somewhat when the time-trend is included. Nevertheless, the qualitative results of the analysis in Section 5 are robust to the choice of specification.

We informally validate the restricted model (with $e_b = e_c$, $e_o = e_p$) by comparing the generated concentrate consumption and pre-allocated fodder quantities with the results from a number of micro-level studies conducted over the past three decades. Table 2 compares the values from various field studies conducted in the 1960s and over the past two decades with the range of model results from 1961-71 and 1976-86 respectively. The comparison between the earlier studies and the model results is quite straightforward and we find that the values obtained are consistent. However, the only comparable recent data are from the Mergos and Slade (1987) and Alderman (1987) studies. Unfortunately, they do not distinguish between cows and buffaloes, providing quantities for milch animals (bovines) in general. Nevertheless, we observe that the fodder and concentrate quantities from these studies lie between the corresponding model values for cow and buffalo.

While the estimated structural coefficients and the generated concentrate quantities appear to be reasonable, formal testing of the over-identifying restrictions still remains. We test the over-identifying restrictions in stages. Our first test is of the most basic restriction $[\pi_4 = \Sigma_i \pi_{5,i} + \pi_{6,i}]$ in equation (6c). This restriction follows from assumptions A2 and A3, given the Cobb-Douglas specification. The respective Wald statistics, distributed as $\chi^2_1$, are 0.570 for cows and 0.084 for buffaloes. Intuitively, the smaller the Wald statistic, the greater the confidence in the null hypothesis since this implies a smaller discrepancy vector. The result of this test supports the basic specification of the model and some of the most crucial assumptions in the derivation of the reduced form.
At the next level we include the assumption that the nutrients from a given feed are equally accessible nutritionally to young and mature animals. As we mentioned earlier, this assumption was implied in the specification of equations (4) and (5). With regard to the reduced form specification, the following restrictions can be derived from equation (6c), \( \pi_2/\pi_3 = \pi_{5,1}/\pi_{6,1}, \forall i \). We test the joint hypothesis of the basic restriction and the four additional restrictions. The three principal testing procedures, the Wald, likelihood ratio, and Lagrange multiplier tests, are only asymptotically equivalent in non-linear models. Since some of our restrictions are now non-linear, we perform multiple tests to ensure that the results are qualitatively consistent. We report results for the likelihood ratio and the Lagrange multiplier tests only (both statistics are distributed as \( \chi^2 \)) since the Wald statistic cannot be conveniently computed when the restrictions are non-linear. The Lagrange multiplier statistic is obtained as 19.974 for cows and 10.221 for buffaloes. The likelihood ratio statistic is obtained as 38.011 and 12.985, respectively. The statistics exceed the critical value in all cases.

As our final exercise, we test all the over-identifying restrictions in the model. The remaining two independent restrictions arise from the equating of concentrate feed elasticities. As expected the null hypothesis is rejected in all cases, with the Lagrange multiplier statistic (distributed as \( \chi^2_k \)) being obtained as 20.083 for cows and 17.615 for buffaloes. The corresponding statistics for the likelihood ratio test are obtained as 38.489 and 29.423, respectively.

The rejection of the over-identifying restrictions required to obtain sensible point estimates implies that neither the restricted nor the unrestricted model unambiguously dominates the other in our case. Consequently, we present results from both models in the subsequent analysis. As we shall see, the qualitative conclusions of our paper are unaffected by the choice of specification.

## 5. The Results of the Model and Implications for Dairy Development

The growth in milk yield in our model occurs either due to increased feeding or due to technological progress in milk production. The objective of the analysis in this section is to derive the respective shares of the growth in milk production due to these two factors. The shares can be directly derived by differentiating equation (6c), after including the quadratic term (number of cooperatives, squared) that we described in the previous section. We compute the share of
production growth due to technological progress, henceforth referred to as TECHSHARE, since the coefficients associated with the number of cooperative societies are relatively stable across the various model specifications.

The share is derived from equation (6c) as,

$$\text{TECHSHARE}_t = \frac{\pi_{\text{coop}} + 2\pi_{\text{coop}^2}\text{COOP}_j\Delta\text{COOP}_j}{\Delta\text{YLD}_t / \text{YLD}_t}$$

(9)

where

- \(\text{TECHSHARE}_t\) = share of milk production growth due to technological progress in year \(t\)
- \(\pi_{\text{coop}}\) = reduced form coefficient corresponding to the number of cooperatives
- \(\pi_{\text{coop}^2}\) = reduced form coefficient corresponding to the number of cooperatives, squared
- \(\text{COOP}_t\) = number of cooperatives in year \(t\)
- \(\text{YLD}_t\) = estimated milk yield in year \(t\)
- \(\Delta\text{COOP}_t\) = change in number of cooperatives (\(\text{COOP}_t - \text{COOP}_{t-1}\))
- \(\Delta\text{YLD}_t\) = change in estimated yield (\(\text{YLD}_t - \text{YLD}_{t-1}\)).

The reduced form coefficients (\(\pi_{\text{coop}}, \pi_{\text{coop}^2}\)) can be directly obtained from the unrestricted OLS estimates. Unfortunately, many of the unrestricted coefficients have perverse signs and unrealistic magnitudes, possibly due to multicollinearity. Alternatively, the reduced form coefficients can be derived from the estimated structural parameters in the restricted model. While the structural parameters have sensible signs and magnitudes, we were forced to reject some of the over-identifying restrictions earlier. Since neither model unambiguously dominates the other, TECHSHARE for cow and buffalo is presented in Table 3 below for each of the models - restricted (NLLS) and unrestricted (OLS). As an additional exercise we estimate the model without young feed terms (\(F_y\) and lagged concentrate-feed prices). The use of lagged values of explanatory variables as independent determinants often tends to exacerbate the multicollinearity problem. Our partially restricted specification attempts to avoid this problem, at the cost of possible omitted variable bias. TECHSHARE obtained from this partially restricted model (ROLS) is also presented in Table 3.

We compute average annual shares over three time periods. The choice of cut-off year for the first time period, beginning in 1962, is obvious since 1961-67 comprises the pre-cooperative regime in our model. The cut-off year for the second period is not as straightforward. We
compute average shares for a number of cut-off years around 1980, since dairy development activity expanded substantially around that time with the inception of Operation Flood II. However, as we see below, our qualitative conclusions are unaffected by the particular choice of the cut-off year for the second period.

The most striking result from Table 3 is the importance of technological progress in increasing milk production. TECHSHARE is generally large in magnitude and, with the exception of buffaloes in the restricted model, increasing over time. TECHSHARE however, varies quite widely over the alternative specifications. Our own preference lies with the restricted model (NLLS) since the point estimates at least are economically sensible. It is also possible to compute standard errors for TECHSHARE from equation (9). TECHSHARE for the restricted model is estimated with far greater precision than it is for the competing specifications (OLS and ROLS). In any case, regardless of the specification that one chooses to believe, TECHSHARE exceeds 20% in period 3 in all cases. This strongly suggests that the cooperative system has played an important role in increasing milk production in recent years.

We provide additional support for our choice of the NLLS specification by deriving the share of the growth in milk production due to the introduction of cross-bred cows (CSHR).\textsuperscript{29} CSHR far

\textsuperscript{29} Given the number of cross-bred and indigenous cows, it is easy to derive the annual increase in milk production due to cross-bred cows (CSHR) if we assume a fixed cross-bred to indigenous cow milk yield ratio over time. The respective average yields are computed as the product of the lactating yield (NC\textsubscript{AER}, 1990) and the proportion of lactating animals (GOI, 1982). The ratio of the average yields is obtained as 3.7221.

\[ YLD_{t} = (YLDC_{t} + YLDD_{t})(NC_{t} + ND_{t}) \]

where,

\[ YLD_{t} = \text{average cow yield in year } t \]
\[ YLDC_{t} = \text{average yield for cross-bred cows in year } t \]
\[ YLDD_{t} = \text{average yield for indigenous cows in year } t \]
\[ NC_{t} = \text{number of cross-bred cows in year } t \]
\[ ND_{t} = \text{number of indigenous cows in year } t. \]

Given \[ YLDC_{t}/YLDD_{t} = 3.7221 \ \forall t, \]
\[ YLD_{t} = (3.7221\cdot NC_{t} + ND_{t})\cdot YLDD_{t}/(NC_{t} + ND_{t}) \]

Alternatively,

\[ YLD'_{t} = (3.7221\cdot NC_{t} + ND_{t} - NC_{t})\cdot YLDD_{t}/(NC_{t} + ND_{t}) \]

where, \[ YLD'_{t} = \text{average cow milk yield, given previous year's number of cross-bred cows.} \]

We finally obtain,
exceeds TECHSHARE in Table 3, for periods 2 and 3, for both the OLS as well as the ROLS specifications. This result is inconsistent with our model since TECHSHARE is specified to include all non-feed inputs, including the effect of cross-breeding. While we do make strong assumptions in deriving CSHR, the order of magnitude of this discrepancy is large enough to suggest that the OLS and ROLS point estimates for TECHSHARE significantly underestimate the importance of technological progress.

A comparison of CSHR with TECHSHARE for the NLLS specification however, leads to the conclusion that direct interventions in the dairy sector cannot explain the growth in cow milk production. The AI program is the most visible scheme associated with the cooperative system and deriving the general magnitude of its impact on milk production allows us to place the other dairy development interventions in perspective. It seems unlikely that these interventions, taken together, could explain the difference between TECHSHARE and CSHR for cows and the large value of TECHSHARE for buffaloes. We consequently conjecture a more diffuse role for the cooperatives in dairy development. They may facilitate a learning process by serving as a channel for the dissemination of information, as well as provide an infrastructure base for the adoption of new technology.

Evidence from the Livestock Census Reports (GOI, 1961, 1966, 1972, 1977, 1982) supports this hypothesis. An increase in the proportion of lactating animals in the herd generally signifies improved animal husbandry and herd management (Shah, 1987) and is by our definition, a measure of technological progress. Results from the Livestock Census Reports suggest that the proportion of lactating indigenous cows has been increasing over time. This result has encouraging distributional implications since indigenous cow owners are generally poorer households compared to buffalo or cross-bred cow owners. It is also consistent with our hypothesis that a broad learning process, possibly associated with the cooperative system, has played an important role in dairy development by improving production techniques. Finally, field studies conducted by the Institute of Rural Management, Anand (IRMA) report that dairy

\[
\text{CSHR} = \frac{(\text{YLD}_t^e - \text{YLD}'_t^e) / (\text{YLD}_t^e - \text{YLD}'_t)}{2.7221(\text{NC}_t - \text{NC}'_t) / ((3.7221 \times (\text{NC}_t + \text{ND}_t))(1 - \text{YLD}_t^e / \text{YLD}_t))}
\]

where, CSHR = share of growth in cow milk yield attributed to the introduction of cross-breeds.
development is observed to occur in well-defined stages that may be associated with a learning process; from subsistence dairying to modern techniques that include cross-breeding (Shah, 1987). While the indirect evidence does suggest the importance of diffuse determinants in dairy development, much more research is clearly required to better understand the role of the cooperatives in this process. This is potentially an important area for future research and has implications for dairy development strategy as well. In this regard, policies that focus directly on the role of the cooperative system as a catalyst for a diffuse development process may be more effective than existing strategies that concentrate on the provision of technical inputs.

6. Conclusion
We develop and estimate a milk supply model for India to derive the sources of growth in milk production. Milk supply in our model is consistent with the individual livestock owner’s profit-maximizing behavior and aggregate fodder availability. The estimated feed elasticities, generated concentrate quantities, and pre-allocated fodder quantities are also consistent with the theory and the corresponding values from various field studies.

Since herd-size is exogenous, milk yield is the dependent variable in our model. The growth in milk yield is stipulated to occur either due to technological progress, associated with the cooperative system, or the increased use of feed. The results of our analysis suggest that a substantial share of this growth in recent years may be attributed to technological progress and that this share has actually been increasing over time. Furthermore, the introduction of cross-bred cows can only account for a fraction of the increase in cow milk yield due to technological progress. This implies an improvement in the quality of indigenous cows as well.

While there appears to have been substantial technological progress in milk production, the directed development initiatives in the dairy sector have generally failed to satisfy their stated objectives. We hypothesize that diffuse determinants play an important role in dairy development. In this regard, the cooperative system may serve as a channel for the dissemination of information, facilitating a broad learning process in the industry, as well as provide an infrastructure base for the adoption of new technology. While we are unable to test our hypothesis, it is consistent with indirect evidence from the dairy sector. A more detailed understanding of this development process, possibly through village-level studies, is a potentially
fruitful area for future research. This hypothesis also has important policy implications. In this regard, policies that focus on the cooperative system's role as a catalyst in dairy development may be more effective than existing strategies that concentrate on the provision of technical inputs.

At the present stage, the model allows only a partial analysis of dairy policy. A milk demand component is required to determine the effect of development interventions on income distribution and nutrition levels. Prices and the herd dynamic should also be endogenous to realistically simulate alternative policy scenarios. Consequently, the model will be integrated with an existing applied general equilibrium model in the future.30

30 See Narayana et. al.(1991) for a description of that model.
References


Amble, V.N., V.V.R. Murty, K.V. Sathe, and B.B.P.S. Goel, 1965, Milk Production of Bovines in India and Their Feed Availability, Indian Journal of Veterinary Science, September, 221-238.

Chandhok, H.L. and The Policy Group, 1990, India Database (New Delhi).


FAO (Food and Agricultural Organization of the United Nations), 1990, Agricultural Supply Utilization Accounts (Rome).

GOI (Government of India), 1976, Report of the National Commission on Agriculture, Part VII (Ministry of Agriculture and Irrigation, New Delhi).


Lau, Lawrence J. and Pan A. Yoipoulos, Profit, Supply and Factor Demand Functions, American Journal of Agricultural Economics, February, 11-18.


Sankhayan, P.L. and A.S. Joshi, 1975, Resource Productivity in Milk Production of Cross-Bred and Indigenous Cows in Rural Areas of Ludhiana District (Punjab), Indian Journal of Agricultural Economics, July-September, 105-111.


Singh, Parmatma and Dayanatha Jha, 1975, Economic Optima in Milk Production, Indian Journal of Agricultural Economics, July-September, 96-105.

### Table 1: Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>COW</th>
<th></th>
<th></th>
<th></th>
<th>BUFALO</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e_b, c, e_o, e_p$</td>
<td>$e_b = e_c, e_o = e_p$</td>
<td>$e_b = e_o, e_p$</td>
<td></td>
<td>$e_b, c, e_o, e_p$</td>
<td>$e_o = e_c, e_o = e_p$</td>
<td>$e_b = e_c, e_o = e_p$</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.260</td>
<td>0.202</td>
<td>-1.584</td>
<td></td>
<td>3.462</td>
<td>0.307</td>
<td>0.253</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.153)</td>
<td>(0.834)</td>
<td></td>
<td>(1.087)</td>
<td>(0.113)</td>
<td>(0.103)</td>
<td></td>
</tr>
<tr>
<td>$e_b$</td>
<td>0.087</td>
<td>0.034</td>
<td>-0.141</td>
<td></td>
<td>0.007</td>
<td>0.058</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.097)</td>
<td>(0.053)</td>
<td></td>
<td>(0.025)</td>
<td>(0.047)</td>
<td>(0.043)</td>
<td></td>
</tr>
<tr>
<td>$e_c$</td>
<td>0.040</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td>0.060</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.110)</td>
<td></td>
<td></td>
<td></td>
<td>(0.049)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_o$</td>
<td>0.021</td>
<td>0.087</td>
<td>0.136</td>
<td>-0.094</td>
<td>0.032</td>
<td>0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.057)</td>
<td>(0.053)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_p$</td>
<td>0.087</td>
<td>-----</td>
<td>-----</td>
<td>-0.001</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td></td>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.018</td>
<td>0.018</td>
<td>0.034</td>
<td>0.011</td>
<td>0.010</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.111E-03</td>
<td>-0.111E-03</td>
<td>-0.254E-03</td>
<td>-0.054E-03</td>
<td>-0.048E-03</td>
<td>-0.033E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.046E-03)</td>
<td>(0.047E-03)</td>
<td>(0.107E-03)</td>
<td>(0.044E-03)</td>
<td>(0.024E-03)</td>
<td>(0.028E-03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>-1.674</td>
<td>-1.519</td>
<td>-1.107</td>
<td>-2.224</td>
<td>-1.363</td>
<td>-1.383</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.449)</td>
<td>(0.423)</td>
<td>(0.133)</td>
<td>(0.289)</td>
<td>(0.376)</td>
<td>(0.385)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.020</td>
<td>0.001</td>
<td>0.036</td>
<td>0.050</td>
<td>0.179</td>
<td>0.192</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.064)</td>
<td>(0.040)</td>
<td>(0.064)</td>
<td>(0.045)</td>
<td>(0.057)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>-----</td>
<td>-----</td>
<td>-0.014</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>D-W</td>
<td>1.398</td>
<td>1.309</td>
<td>1.507</td>
<td>1.909</td>
<td>2.087</td>
<td>2.059</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.961</td>
<td>0.960</td>
<td>0.971</td>
<td>0.973</td>
<td>0.949</td>
<td>0.949</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
Standard errors are in parentheses

- $a$ = young-feed elasticity multiplier
- $e_b$ = bran elasticity
- $e_c$ = coarse grains elasticity
- $e_o$ = oilseeds elasticity
- $e_p$ = pulses elasticity
- $\lambda_1$ = (number of cooperatives, in thousands) coefficient
- $\lambda_2$ = (number of cooperatives, squared) coefficient
- $\pi_1$ = constant term (reduced form)
- $\theta$ = (dummy variable, 1961-67) coefficient
- $t$ = time-trend
- $n$ = number of observations
Table 2: Comparison between Micro-study and Model Quantities

<table>
<thead>
<tr>
<th>EARLY STUDIES</th>
<th>Cow (Kg./animal/day)</th>
<th>Buffalo (Kg./animal/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dm</td>
<td>Gm</td>
</tr>
<tr>
<td>Model (1961-71)</td>
<td>3.43-3.78</td>
<td>2.46-2.56</td>
</tr>
<tr>
<td>Karnataka (1965-66)</td>
<td>4.14</td>
<td>3.81</td>
</tr>
<tr>
<td>Tamil Nadu (1968-69)</td>
<td>6.44</td>
<td>1.67</td>
</tr>
<tr>
<td>A.P. (1971-72)</td>
<td>4.62</td>
<td>1.22</td>
</tr>
<tr>
<td>U.P. (1962-63)</td>
<td>3.17</td>
<td>3.50</td>
</tr>
<tr>
<td>Gujarat (1963-64)</td>
<td>3.17</td>
<td>1.92</td>
</tr>
<tr>
<td>Gujarat (1958-59)</td>
<td>2.91</td>
<td>2.47</td>
</tr>
<tr>
<td>Bihar (1966-67)</td>
<td>3.99</td>
<td>3.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RECENT STUDIES</th>
<th>Bovine: Cow+Buffalo (Kg./animal/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dm</td>
</tr>
<tr>
<td>Model (1976-86)</td>
<td>3.77-4.35</td>
</tr>
<tr>
<td>M.P. (1983)(^a)</td>
<td>2.55</td>
</tr>
<tr>
<td>Karnataka (1984)(^b)</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Note:
Figures in brackets refer to the year in which the data was collected. The quantities listed for the model correspond to the range for the early years (1961-71) and the recent years (1976-86).

All the regional data reported in the early studies are from Vaidyanathan (1988) and are attributed to various studies by Daroga Singh et. al., New Delhi, IARS, ICAR (except for the 1958-59 Gujarat data which is referenced as GOI, IARS, 1969). The data were provided separately for dry and lactating animals. The proportion of lactating animals is obtained for all the regional studies from Daroga Singh et. al., except for Karnataka which is obtained from the Livestock Census (GOI, 1966).

The following studies are not included in the Table:
1. Punjab and Haryana studies, since they generally display substantially higher values than the all-India average for milk yield as well as for input consumption.
2. NCAER survey results (NCAER, 1990), since they are on the high side although derived from a national sample. In fact, the NCAER results are remarkably similar to Sankhayan and Joshi’s Punjab data (Sankhayan and Joshi, 1975) suggesting that the national sample may have been biased.
3. Lalwani’s (1989) results, since these are not consistent with other studies. There may have been problems in data collection or processing in this study as well.

\(^a\) - Mergos and Slade (1987) - milch animals and bullocks are assumed to have the same rations
\(^b\) - Alderman (1987).
Table 3: Decomposition of the Annual Growth in Milk Yield

<table>
<thead>
<tr>
<th></th>
<th>COW</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YLD</td>
<td>NLLS</td>
<td>OLS</td>
<td>ROLS</td>
<td>CSHR</td>
<td>YLD</td>
<td>NLLS</td>
<td>OLS</td>
<td>ROLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1962-67</td>
<td>-1.598</td>
<td>0.265(0.053)</td>
<td>0.010(0.072)</td>
<td>0.125(0.172)</td>
<td>0.003(0.003)</td>
<td>-1.880(0.050)</td>
<td>0.418(0.148)</td>
<td>0.181(0.182)</td>
<td>0.205(0.182)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968-78</td>
<td>2.945</td>
<td>0.317(0.126)</td>
<td>0.010(0.239)</td>
<td>0.163(0.242)</td>
<td>0.238(0.242)</td>
<td>1.889(0.146)</td>
<td>0.708(0.525)</td>
<td>0.286(0.407)</td>
<td>0.449(0.407)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968-79</td>
<td>2.945</td>
<td>0.340(0.142)</td>
<td>0.017(0.289)</td>
<td>0.200(0.313)</td>
<td>0.249(0.313)</td>
<td>1.951(0.146)</td>
<td>0.733(0.525)</td>
<td>0.262(0.407)</td>
<td>0.411(0.407)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968-80</td>
<td>3.069</td>
<td>0.391(0.142)</td>
<td>0.015(0.289)</td>
<td>0.185(0.313)</td>
<td>0.250(0.293)</td>
<td>2.137(0.169)</td>
<td>0.742(0.495)</td>
<td>0.254(0.406)</td>
<td>0.406(0.395)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968-81</td>
<td>3.114</td>
<td>0.401(0.158)</td>
<td>0.016(0.269)</td>
<td>0.178(0.293)</td>
<td>0.258(0.293)</td>
<td>2.242(0.174)</td>
<td>0.727(0.478)</td>
<td>0.250(0.406)</td>
<td>0.406(0.395)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1979-86</td>
<td>5.254</td>
<td>0.781(0.587)</td>
<td>0.210(0.844)</td>
<td>0.337(0.786)</td>
<td>0.410(0.410)</td>
<td>4.177(0.261)</td>
<td>0.699(0.616)</td>
<td>0.324(0.499)</td>
<td>0.500(0.499)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980-86</td>
<td>5.583</td>
<td>0.809(0.661)</td>
<td>0.227(0.885)</td>
<td>0.298(0.624)</td>
<td>0.417(0.624)</td>
<td>4.398(0.261)</td>
<td>0.656(0.616)</td>
<td>0.370(0.499)</td>
<td>0.572(0.499)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981-86</td>
<td>5.753</td>
<td>0.757(0.661)</td>
<td>0.265(0.885)</td>
<td>0.348(0.780)</td>
<td>0.442(0.780)</td>
<td>4.402(0.244)</td>
<td>0.623(0.700)</td>
<td>0.406(0.557)</td>
<td>0.611(0.557)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982-86</td>
<td>6.166</td>
<td>0.825(0.778)</td>
<td>0.313(1.080)</td>
<td>0.400(0.994)</td>
<td>0.458(0.994)</td>
<td>4.563(0.262)</td>
<td>0.641(0.818)</td>
<td>0.449(0.637)</td>
<td>0.652(0.637)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note:
Standard errors are in parentheses
YLD = average annual percentage growth in milk yield
NLLS = TECHSHARE computed with restricted estimates
OLS = TECHSHARE computed with unrestricted estimates
ROLS = TECHSHARE computed without young feed variables
CSHR = share of growth in cow milk yield accounted for by introduction of cross-breeds.

Computed shares exceed one or are negative in some years depending on the relative variation in milk yield and the number of cooperatives. We correct such observations, restricting all annual shares to the range between 0 and 1. Such restricted observations are not included when computing the standard errors.
Chapter 3:
Social Learning and Technology Diffusion: An Application to Indian Agriculture

1. Introduction

The invention of new and superior technology provides no benefit to society until it is adopted by firms in the economy. Technology diffusion, in conjunction with R&D, consequently plays a critical role in economic development. Perhaps in recognition of its importance, the study of technology diffusion has a long tradition in the economics literature. Early neoclassical models studied diffusion as a contagion process in which firms operating in competitive markets adopted the new technology with fixed probability once their neighbors had adopted (Griliches, 1957, and Mansfield, 1961, are classic papers in this literature). More recent research has sought to place a richer structure on the contagion process underlying the neoclassical model. Diffusion is interpreted in this literature as the outcome of a social learning process in which agents (or firms) learn from their neighbors about the quality of a new and uncertain technology (see for instance, Banerjee, 1992, 1993, Bikhchandani, Hirshleifer, and Welch, 1992, Ellison and Fudenberg, 1993). Agents with superior information or favorable endowments adopt the new technology first and are then followed by their neighbors, who need to be assured about the quality of the technology before investing.\(^{31}\) The principle objective of the paper is to gain a better understanding of this social learning process.

There is general agreement in the literature about the motivation for social learning; agents condition their decisions on the decisions and experiences of their neighbors, since these provide information about the new technology. No consensus has emerged, however, with regard to the specific structure that is to be placed on the learning process.\(^{32}\) Alternative models of social learning may be broadly distinguished by the agent's ability to extract

---

\(^{31}\) Strategic interaction among rival firms in concentrated markets, which may also lead to sequential adoption of the new technology, is ignored in this paper. Reinganum, 1989, provides an exhaustive survey of the literature.

\(^{32}\) Examples in the received literature range from models of Bayesian learning, in which agents use all of the available information optimally, to naive models in which firms use simple, exogenously specified, rules-of-thumb for learning. Among the papers referred to earlier, Banerjee, 1992, 1993, and Bikhchandani, Hirshleifer, and Welch, 1992, assume Bayesian learning, whereas Ellison and Fudenberg, 1993, specify exogenous rules-of-thumb.
information from neighbors whose characteristics (types) differ from his own. When learning is efficient, the agent is able to condition for differences in neighbors' types to learn about the new technology. Consequently, all neighbors have equal influence in the agent's learning function, regardless of their distance from him in type space. The conditioning described above uses knowledge of the production process which, if incorrect, would result in biased inference about the quality of the new technology. Consequently, the agent may rationally prefer to use relatively inefficient learning rules that discount valuable information from distant (in type space) neighbors, but place little structure on the production function. The optimal learning rule is determined, in practice, by trading-off efficiency and potential specification bias. A more detailed discussion on the alternative learning rules, in the context of the particular application chosen for the empirical analysis, is provided later in this section.

The choice of learning rule is not an innocuous assumption, since the convergence results of theoretical models are often predicated on the structure of the learning process (Banerjee, 1993). Furthermore, optimal policy, aimed at spreading the new technology, depends crucially on the nature of the learning process. The policy implications associated with alternative learning models assume particular importance in light of the well documented evidence that new technologies, especially in agriculture, generally require many decades to diffuse completely. In such circumstances, the correct choice of policy intervention, which in turn is determined by the underlying learning structure, could accelerate the steady-state adoption of the new technology by many years. A more detailed discussion on optimal policy interventions under alternative learning regimes is postponed until later in the paper.

Determining the structure of the learning process is ultimately an empirical problem. In this case we are faced with a double inference problem; to infer from the agent's observed

---

33 See, for instance, Timmer's (1969) discussion on the slow rate of diffusion of new technology in European agriculture.

34 The optimal policy intervention follows naturally from the structure of the learning process. For instance, when social learning is inefficient, valuable information from distant (in type space) neighbors is discounted during the learning process. Endogenous diffusion of the new technology may then be relatively slow, especially in heterogeneous populations, requiring external interventions such as exogenous signals from government agencies and the media to disseminate information. In contrast, when learning is efficient, external information may be only required to initiate the endogenous learning process.
actions, and the prior actions and outcomes of those around him, the manner by which the agent himself infers. My strategy in this study is to choose a convenient application, derive alternative learning models, and then determine which of these models is most consistent with the observed pattern of technology diffusion. As described above, the implications of the alternative learning models are distinguished by the effect of distance (in type space) on a neighbor’s influence in the learning function.

The adoption of Green Revolution technology in Indian agriculture is chosen for the empirical analysis.35 The new High Yield Variety (HYV) technology for wheat and rice was introduced in India in the 1960s. It is characterized by considerably higher yields, albeit with greater uncertainty in production, when compared with traditional varieties. Over the past three decades, HYV technology has diffused over most of the wheat and rice growing areas of the country. There are two advantages to choosing agriculture as the application in which to study learning. First, agriculture is a competitive industry, allowing one to ignore strategic aspects of the adoption decision. Second, agricultural investment occurs with fixed frequency, allowing us to ignore the timing of investment.

The learning models proposed in this paper are all broadly consistent with individual rationality. The empirical results are consequently relevant to the long-standing debate regarding the extent to which peasant behavior is influenced by social norms (see Behrman, 1966, for a discussion). An observed pattern of spatial diffusion that is consistent with any of the proposed rational learning models would weakly support the position that peasant behavior is not entirely norm-driven, at least with regard to investment decisions.36

---

35 While there are numerous references to the importance of social learning in the adoption of new agricultural technology (see for instance, Birkhauser, Evenson, and Feder, 1991, and Feder, Just, and Zilberman, 1985), there have been few attempts to implement empirical models that explicitly incorporate this effect. Recent empirical studies that use neighborhood effects to explain the adoption of new agricultural technology include, Besley and Case, 1994, Case, 1992, Burger, Collier, and Gunning, 1993, and Foster and Rosenzweig, 1994.

36 Such a result would still be consistent with the case in which individual behavior is jointly determined by social learning and pure social effects, such as norms or peer influence.
The peasant's objective, in my model, is to determine the performance (expected yield) of the new crop at his particular location.\textsuperscript{37} Yield is, in general, a function of locational (geographic) attributes, such as soil quality, and the peasant's characteristics (type). The peasant's type is in this case an index constructed from observed characteristics such as income, irrigated land, size of land-holdings, and asset ownership. Geographically contiguous neighbors provide information about the new technology since their locational attributes are presumably the same as the peasant's. The peasant obtains the best estimate of his expected yield in each period from his neighbors' (and his own) previous-period acreage allocations and yield realizations, as well as from an exogenous signal. Yield realizations, of course, provide a direct estimate of the expected yield. Estimates of the expected yield can also be obtained from the acreage decision since acreage is, in turn, increasing in expected yield. The parametric specification of the acreage function is assumed to be common knowledge throughout the paper.\textsuperscript{38} Finally, acreage allocations and yield realizations provide independent information about the expected yield if we assume that agricultural production exhibits constant returns to scale with respect to total acreage. Given the best estimate of his expected yield, the peasant is able to determine the optimal acreage to be allocated to the new crop in each period.

The peasant's inference problem is complicated by the presence of the "type" term in the yield function. Since the peasant and his neighbors are unlikely to share a common type, biased estimates of his expected yield will be obtained from them. Since this bias is increasing in distance (in type space), information received from different neighbors is discounted in the peasant's learning function, with the result that learning will be inefficient. Alternatively, the peasant can obtain unbiased expected yield estimates from his neighbors, in principle, by placing restrictions on the yield function, which will permit him to condition for differences in

\textsuperscript{37} The peasant requires this information since the acreage allocated to the new technology is increasing in expected yield. Higher moments of the yield function, such as the variance, are assumed to be known in this paper. Future research could relax this assumption.

\textsuperscript{38} The acreage function is endogenously specified by the peasant, in contrast with the yield function which is determined by an unknown law of nature. There is consequently no uncertainty associated with the parametric specification of the acreage function, whereas the true parametric specification of the yield function is never determined with certainty.
their types. While learning will now be efficient, since all neighbors are equally influential in the learning function, such restrictions give rise to a new cost associated with potential misspecification of the yield function. In practice, the peasant will choose the optimal structure to be placed on the yield function by trading-off potential specification bias with efficiency.

The alternative learning models proposed in this paper are distinguished by the amount of structure that is placed on the yield function. Following the econometrician’s standard problem, consider the canonical learning models, with regard to the restrictions placed on the yield function; unrestricted nonparametric (kernel) estimation places no restrictions on the yield function, in contrast with parametric estimation. The implications of the canonical learning models are quite distinct, being distinguished by the effect of distance (in type space) on a neighbor’s influence (weight) in the learning function. Following the discussion above, neighbor-influence is monotonically declining in distance with nonparametric learning. In contrast, all neighbors are equally influential with parametric learning since the strong restrictions on the yield function allow the peasant to control for differences in neighbors’ types. As we shall see later in the paper, the estimated pattern of neighbor-influence is consistent with neither of the canonical learning models. A third model is consequently proposed, ex post, to explain the observed pattern. This model places a single, monotonicity, restriction on the nonparametric model described above. The implications of the intermediate, linear smoothing, model include (possibly) negative weights on distant (in type space) neighbors and a nonmonotonic pattern of learning weights in distance.

The implications of the alternative learning models are tested with district-level data over a 22 year period. Acreage allocations, yield realizations and exogenous characteristics are obtained for each district in each year. A simple two-stage econometric procedure is implemented to estimate the pattern of learning weights. In the first stage, acreage and yield are estimated as functions of the district’s exogenous characteristics in each year. In the second stage, each district’s acreage residual, obtained from the first-stage, is regressed on its geographically contiguous neighbors’ (and its own) previous-period acreage and yield residuals, obtained in

---

39 It is assumed throughout the paper that neighbors’ types can be observed by the peasant.
turn from the first stage. Since neighbors are ranked by their distance from the district (in type space) in the second stage, the pattern of learning weights across neighbors reveals the underlying learning process. The attempt to determine the influence of different neighbors in the learning function, on the basis of their distance in type space, can be distinguished from most previous empirical studies of social learning which treat neighbors as a group.46

The intuition for the two-stage estimation procedure described above is fairly straightforward. The first stage controls for changes in exogenous characteristics that would independently result in the spatial diffusion of the new technology over time. The residuals from the first-stage are consequently interpreted as shocks providing previously unavailable information about the new technology. Consider the effect of a shock to a district’s acreage allocation in a given period. If its neighbors’ subsequent-period shocks are positively correlated with this shock, independently of their distance from the district in type space (rank), then the spatial pattern of residuals over time suggests a model of parametric learning. In contrast, a pattern in which near (in type space) neighbors’ shocks are more highly (positively) correlated with the district’s previous-period shock would follow from unrestricted nonparametric learning.

Two econometric problems are associated with the attempt to infer the underlying learning process from the spatial diffusion of the new technology over time at the district level; the identification problem and the aggregation problem. The identification problem arises when both social learning and unobserved correlated effects occur simultaneously. Under such conditions it is impossible to distinguish social learning, in which individual decisions are conditioned on neighbors’ decisions, from correlated effects, in which neighboring individuals receive correlated, unobserved, shocks (see Manski, 1993b, for

---

46 Such studies implicitly place equal weight on all neighbors in the learning function. See for instance, Case, 1992, Besley and Case, 1994, and Foster and Rosenzweig, 1994. In contrast, Burger, Collier, and Gunning, 1993, partition the set of neighbors by sex of household-head. This partitioning is, however, not fine enough to distinguish between alternative learning models.
this result).\footnote{Manski, 1993b} We avoid this problem because learning occurs with a single-period lag in agriculture, whereas correlated effects are presumably contemporaneous.

The aggregation problem follows from our use of aggregate, district-level, data to test for disaggregate, farm-level, learning. The following simple example illustrates the problem, although the result obtained is more general. Suppose that a peasant in a far corner of a district receives an exogenous signal about the new technology and increases the acreage allocated to it, in a given period. Suppose, further, that this shock does not propagate rapidly enough through the system to reach the adjacent district by the next period (planting). A single-period aggregate learning rule will fail to pick up the shock in this case, underestimating the learning effect. One solution to the dilution problem is to include multiple lags at the aggregate level, even when disaggregate learning entails conditioning on the previous period only, since shocks will eventually propagate through the system. This result permits us to test for the presence of aggregation problems; aggregate learning will appear to follow a multi-period rule even when disaggregate learning is characterized by a single period rule. Empirical evidence presented later in the paper suggests that aggregation does not create a severe problem in our case.

As discussed earlier, the estimation results reveal a pattern of neighbor-influence that is broadly consistent with the intermediate, linear smoothing, model of learning. While any inference about the specific structure of the learning process should be treated cautiously, this result generally supports the hypothesis that social learning is fairly efficient. Peasants appear to condition, albeit crudely, for differences in neighbors' types when learning from them. This result suggests that peasant behavior may not be entirely norm-driven, at least with regard to investment decisions. In addition, it implies that learning models that ignore heterogeneity, and treat neighbors as a group, may be misspecified since estimated learning weights vary across neighbors (by distance in type space). With regard to policy implications, information will endogenously diffuse quite rapidly, even through a heterogeneous population.

\footnote{Manski, 1993b, also considers exogenous social effects, in which agents condition their decisions on their neighbors' exogenous characteristics. This effect is typically associated with technological or pecuniary externalities in production, or with strategic behavior in concentrated markets. It can be ignored for our application since agriculture is a competitive industry in which non-informational externalities are absent.}
with such a learning model. Consequently, policy interventions that facilitate the social learning process may be as effective as exogenous information signals in spreading the new technology.

The paper is organized in five sections. Section 2 describes the peasant's inference problem. Section 3 describes the econometrician's inference problem; how she might infer, from the observed spatial pattern of acreage allocations and yield realizations, how the peasant himself inferred. Section 4 discusses the results of the estimation procedure and Section 5 concludes the paper.

2. The peasant's inference problem

As discussed in the previous section, the peasant's objective is to determine the expected yield corresponding to his location, in type space and geographical space. He computes his best estimate of the expected yield in each period using information obtained from his neighbors' (and his own) previous-period acreage allocations and yield realizations. The canonical learning models differ in the amount of structure that is placed on the yield function. Nonparametric kernel estimation places no structure on the yield function. In contrast, parametric learning places strong parametric restrictions on the yield function. Finally, the linear smoothing model that is proposed to explain the observed pattern of learning weights, places a single monotonicity restriction on the yield function; yield is increasing in the peasant's type. Learning functions corresponding to these alternative models will be derived below.

To begin with, consider the following alternative specifications of the yield function. The first specification in equation (1) places no parametric restrictions on the yield function; yield is simply some function of the peasant's type. This contrasts with the second specification which places strong, parametric, restrictions on the yield function.

\[ y_{it} = m(z_{it}) + \eta_{it} = f(z_{it}; \theta, \beta) + \eta_{it} \]  

(1)

where, \( y_{it} \) = yield (production per acre) for peasant, \( i \), in period, \( t \).
$z_a$ = index that reflects the peasant’s type, which is in turn determined by a vector of exogenous characteristics. It is not necessary to specify a unidimensional function here. This will, however, simplify the exposition of the alternative learning models significantly.

$\eta_a$ = stochastic term associated with uncertainty in the yield, $\eta_a \sim N(0, \lambda^2)$. 

$\beta$ = elasticity of yield with respect to type. This is assumed to be known, for simplicity. This assumption will later be relaxed.

$\theta$ = parameter representing locational characteristics. The true value of this parameter is not known to the peasant.

With unrestricted nonparametric learning the peasant uses the first yield specification in equation (1), without any restrictions, to determine the expected yield corresponding to his geographical location and his type, $m(z_a)$. In contrast, with parametric learning the peasant uses the second specification in equation (1), with parametric restrictions, to determine his expected yield. Finally, the linear smoothing model places a single, nonparametric, restriction on the yield function, $m'(z_a) > 0$.

Since we assume for simplicity that the elasticity, $\beta$, is known, the peasant’s immediate objective with parametric learning is to determine the value of the locational parameter, $\theta$. Once an estimate of $\theta$ is obtained, it is straightforward for the peasant to compute his expected yield in each period. The locational parameter, $\theta$, is common across neighbors, in contrast with the expected yield which varies with type. This feature of parametric learning permits the peasant to efficiently extract all the relevant information about the new technology from his neighbors. Nonparametric learning, in contrast, discounts information received from distant (in type space) neighbors, and so is less efficient. As discussed previously, the peasant may nevertheless rationally prefer nonparametric learning to the more efficient, but less robust, parametric approach when he is suitably uncertain about the true specification of the yield function.

Turning next to the acreage function, recall from our previous discussion that the parametric specification of the acreage function is known to the peasant with certainty since it is endogenously specified by him. An estimate of the expected yield can consequently be obtained from the acreage allocation decision since acreage is, in turn, increasing in expected yield. The remaining arguments of the acreage function are assumed to be known in this case.
Acreage specifications corresponding to the alternative yield specifications discussed earlier are presented below,

\[ A_{ii} = g(m^A(z_{ui}), R_i, P_i, \tau_{ui}^A, \gamma_i; \beta) = g(z_{ui}, \theta_{ui}^A, R_i, P_i, \tau_{ui}^A, \gamma_i, \beta) \quad A_{ii} \geq 0 \]  

(2)

where, \( A_{ii} \) = acreage allocated to the new crop by peasant, \( i \), in period, \( t \).
\( m^A(z_{ui}) \) = peasant’s best estimate of the expected yield, \( m(z_{ui}) \), in period, \( t \).
\( \theta_{ui}^A \) = peasant’s best estimate of the locational parameter, \( \theta \), in period, \( t \).
\( R_i \) = risk-aversion coefficient.
\( P_i \) = input and output prices.
\( \tau_{ui}^A \) = precision in \( m^A(z_{ui}) \) or \( \theta_{ui}^A \) estimate. The precision is defined as the reciprocal of the variance. This term reflects the peasant’s confidence in his current assessment of the new technology.
\( \tau_{ui}^y = 1/\lambda_{ui}^2 \) = yield precision.
\( \gamma \) = acreage function parameters, which are assumed to be known.

The first specification in equation (2) corresponds to nonparametric learning, in which the peasant’s objective is to directly estimate his expected yield, \( m(z_{ui}) \), in each period. The second specification in equation (2) corresponds to parametric learning in which the peasant estimates the locational parameter, \( \theta \), in each period. Equation (2) can be obtained as the outcome of a simple portfolio-choice problem under uncertainty. Risk-aversion generates interior solutions for the acreage allocation, \( A_{ii} \), in this case.\(^{42}\)

\(^{42}\) For instance, suppose that two technologies are available; the new, risky, technology and a traditional, safe, technology. The traditional technology provides a certain output with zero net return. Assume CARA preferences and normally distributed payoffs. For the time being, assume that the expected yield is known with certainty and ignore prices. The peasant’s problem is,

\[ \max EU(W_{it}) = \max R_i E(W_{it}) - \frac{1}{2} R_i^2 \text{var}(W_{it}) \]

\[ W_{it} = A_{ii} y_{it}, \quad y_{it} = m(z_{ui}) + \eta_{it}, \quad \eta_{it} \sim N(0, \lambda_{it}^2) \]

where, \( W_{it} \) = peasant’s wealth in period, \( t \).
\( R_i \) = coefficient of absolute risk-aversion.

Maximizing expected utility, \( EU(W_{it}) \), with respect to acreage, \( A_{ii} \), the optimal acreage allocation is derived from the first-order condition as,

\[ A_{ii} = \frac{m(z_{ui})}{R_i \lambda_{it}^2} \]

The additional arguments in the acreage function in equation (2) are obtained by introducing prices in the peasant’s optimization problem and by considering uncertainty in the expected yield, \( m(z_{ui}) \).
The issue of experimentation by peasants, with a consequent forward-looking component to the acreage allocation decision, is ignored in this paper primarily to reduce the complexity of the model. This simplification may not be entirely unreasonable for our agricultural application. Information about the new technology is available from multiple neighbors, reducing the relative value of individual experimentation. Recent econometric evidence from farm-level Indian agricultural data, however, suggests that experimentation may be an empirically significant determinant of the acreage allocation decision (Besley and Case, 1994, Foster and Rosenzweig, 1994). While experimentation is not the primary focus of this paper, I will nevertheless attempt to control for its presence, in the estimation procedure.

In addition to information from neighbors’ acreage allocations and yield realizations, the peasant also receives exogenous signals about the new technology from the local extension worker, and other information sources, in each period. These signals are assumed to provide unbiased estimates of \( m(z_a) \) or \( \theta \), depending on the learning model under consideration.

Learning functions corresponding to the alternative models are derived next. Under the conditions described below, the peasant’s best estimate of \( m(z_a) \), or \( \theta \), is specified as a linear function of the corresponding estimates obtained from his neighbors’ (and his own) previous-period acreage allocations and yield realizations, as well as the estimate received from the exogenous signal.

**Learning Model 1: Parametric estimation**

This model places parametric restrictions on the yield function. To simplify the exposition I assume that the type elasticity, \( \beta \) in equation (1), is known. Consequently only a locational parameter, \( \theta \), remains to be determined. The peasant’s objective in this model is to learn the value of \( \theta \). Since this parameter is common across neighbors, all neighbors have equal influence in the learning function under fairly weak conditions that I will discuss below. The peasant is consequently able to condition for differences in neighbors’ types to derive efficient estimates of \( \theta \), assuming of course that the parametric restrictions are correctly chosen.\(^3\)

\(^3\) The assumption that \( \theta \) is time-stationary is quite strong. In practice, successive generations of the HYV crop display varying characteristics, as they mix with traditional varieties over time. While the time-stationarity assumption may over-state the influence of neighbors’ prior decisions and experiences, since previous-period \( \theta \)
Estimates of $\theta$ are obtained from neighbors’ yield realizations and acreage allocations, following the parametric specifications described in equations (1) and (2). For this we also require that the yield function be invertible, in the sense that an unbiased, normally distributed, estimate of $\theta$ can be obtained from a single observation.\footnote{Recall that the stochastic term in the yield function was assumed to be normally distributed in equation (1). Assuming a single regressor, the following specifications for instance are acceptable,}

$$y_u = \theta + \beta z_u + \eta_u \rightarrow \hat{\theta} = (y_u - \beta z_u)$$

$$y_u = \theta z_u + \eta_u \rightarrow \hat{\theta} = y_u / z_u$$

We obtain an unbiased estimate of $\theta$ in both cases if the usual orthogonality conditions are satisfied. In contrast,

$$y_u = z_u + \eta_u$$

does not provide an unbiased, normally distributed, estimate of $\theta$ with a single observation.

\footnote{In this case the maximum likelihood estimate of $\theta$ is consistent with Bayesian inference since the peasant’s prior, $\theta^{\text{a},i}$, enters the likelihood function. The likelihood function is derived by assuming that all the estimates of $\theta$ obtained from the three sources of information are drawn from independent normal distributions with mean $\theta$. The log-likelihood function is,}

$$L = -\frac{2(n+1)}{2} \ln(2\pi) - \frac{1}{2} \sum_{j \in L_i} \left( \ln \sigma_{\theta_{j-1}}^2 + \ln \lambda_{\theta_{j-1}}^2 \right) - \frac{1}{2} \ln \nu^2 - \frac{1}{2} \sum_{j \in L_i} \left[ \frac{1}{\sigma_{\theta_{j-1}}} (\theta_{\theta_{j-1}} - \theta)^2 + \frac{1}{\lambda_{\theta_{j-1}}^2} (\theta_{\theta_{j-1}} - \theta)^2 \right] - \frac{1}{2} \frac{1}{\nu^2} (\theta - \theta)^2$$

where, $n =$ size of learning set, $L_i$. Maximizing $L$ with respect to $\theta$, we obtain $\theta^{\text{a},i}$ as a precision-weighted average of $\theta_{j-1}^{\text{a},i}$, $\theta_{j-1}^{\text{a},j}$, $\forall j \in L_i$, $\theta^{\text{a},i}$. This implies, $\theta^{\text{a},i} \sim N(\theta, \sigma_{\theta}^2)$. Note that $i \in L_i$.

The assumption that neighbors’ $\theta$ estimates are drawn from independent distributions is very strong. Some of the peasant’s neighbors are likely to be neighbors to each other as well, consequently providing correlated signals about the new technology. In such cases, a rational peasant would place less weight on such neighbors’ signals to avoid double-counting information. Similarly, the peasant should account for the fact that a neighbor’s previous-period signal is based, in part, on information received from the peasant himself, two periods back. His prior, $\theta^{\text{a},i}$, will consequently be correlated with neighbors’ previous-period signals, $\theta^{\text{a},j}$, $j \neq i$. However, the bias that follows by ignoring these correlations is unlikely to qualitatively affect the basic implication of parametric learning; learning
\[
\theta_d^t = \frac{\sum_{j \in L_t} (\tau^t_{j-1} - \theta^t_{j-1} + \tau^t_{d, -1} + \tau^t_{j, -1}) + \tau^t_{d}}{\sum_{j \in L_t} (\tau^t_{j-1} + \tau^t_{j, -1}) + \tau^t_{d}}
\]

where, \(\theta^t_{j, -1}, \theta^t_{d, -1}\) = estimates of \(\theta\) obtained from neighbors' previous-period acreage allocations and yield realizations, respectively.
\(\theta^t_{d}\) = estimate of \(\theta\) obtained from the exogenous signal.
\(\tau^t_{j, -1}, \tau^t_{d, -1}\) = acreage and yield precisions. Recall that the precision is defined as the reciprocal of the variance.
\(\tau^t_{d} = 1/\nu^t_{d}\) = precision of the exogenous signal.
\(L_t\) = peasant \(i\)'s learning set. This includes the peasant himself.

The variance of \(\theta^t_{d}\) can be easily obtained if we assume that \(\tau^t_{j, -1}, \tau^t_{d, -1}, \tau^t_{d}\) are known constants.

\[
\tau^t_d = \frac{1}{\text{var}(\theta^t_d)} = \sum_{j \in L_t} (\tau^t_{j-1} + \tau^t_{d, -1}) + \tau^t_{d}
\]

Substituting from equation (4) in equation (3), the learning function is expressed as,

\[
\theta_d^t = \sum_{j \in L_t} (w^j_{\lambda} \theta^t_{j, -1} + w^j_{\nu} \theta^t_{d, -1}) + w^t \theta^t_{d}
\]

\[
w^j_{\lambda} = \begin{bmatrix} \tau^t_{j-1} \\ \tau^t_d \end{bmatrix}, \quad w^j_{\nu} = \begin{bmatrix} \tau^t_{j-1} \\ \tau^t_d \end{bmatrix}, \quad w^t = \begin{bmatrix} \tau^t_d \\ \tau^t_d \end{bmatrix}
\]

\(w^t_{\lambda}, w^t_{\nu}, w^t\) = learning weights placed on \(\theta\) estimates obtained from neighbors' acreage decisions, neighbors' yield realizations, and the exogenous signal, respectively.

Time-stationarity of the learning weights is assumed here, and in the subsequent learning models, for simplicity. Since \(\theta^t_{j, -1} \sim N(\theta^t, \lambda^2_{j-1}), \theta^t_{d, -1} \sim N(\theta, \nu^2_d),\) this implies \(\theta^t_{d} \sim N(\theta, \sigma^2_d).\) No one is systematically misinformed about the locational characteristics of the new technology in the parametric model. Suppose now that neighbors are ranked by distance, in type space, in

weights continue to be independent of neighbors' distance from the peasant (in type space) as long as the bias is uncorrelated with distance.
equation (5). It is reasonable to expect that high types provide more precise information about the new technology. However, high and low types are treated symmetrically in the learning model; high types learn from low types, and vice versa. Consequently, a high type is as likely to be a distant neighbor as a low type and no correlation between signal precision and neighbor rank will be observed in general. All neighbors have equal influence in the parametric learning model, regardless of their rank.\textsuperscript{46}

\textbf{Learning Model 2: Unrestricted nonparametric (kernel) estimation}

In this model, the peasant's objective is to directly estimate his expected yield in each period. As discussed previously, the difficulty with nonparametric estimation arises when neighbors' types differ from that of the peasant, since he does not control for such differences when learning from them. The bias associated with the use of a neighbor's expected yield estimate, as an estimate of the peasant's own expected yield, is increasing in the neighbor's distance (in type space) from him. Since the peasant must trade-off efficiency with bias in learning from his neighbors, distant (in type space) neighbors are consequently less influential in the nonparametric learning function. This contrasts with the implication of the parametric model derived above. Neighbors have equal influence in that learning model since they all provide unbiased estimates of the locational parameter, \( \theta \), regardless of their distance from the peasant in type space.

The peasant's expected yield is specified as a weighted average of the various expected yield estimates that he obtains from his neighbors' previous-period acreage allocations and yield realizations, as well as from the exogenous signal.\textsuperscript{47}

\textsuperscript{46} The implications of the alternative learning models are derived in terms of a neighbor's influence, as a function of his rank from the peasant in type space. Neighbors are consequently sorted by rank in the estimation procedure, and the pattern of estimated coefficients (by rank) permits inference about the structure of the underlying learning process. The use of ranks allows us to obtain inferences that are relatively robust to the econometrician's choice of the distance metric.

\textsuperscript{47} The specification of a single-period learning rule is not strictly correct in this case. Consider, for instance, a simple learning rule in which neighbors beyond some cut-off distance (in type space) are excluded from the learning function. Suppose a neighbor's period-0 type lies beyond the cut-off, excluding his signal from the peasant's period-1 learning function. If the peasant's type subsequently changes sufficiently, the neighbor's period-0 type could conceivably lie within the period-2 cut-off. The neighbor's period-0 signal would then enter the peasant's period-2 learning function, implying a multi-period learning rule. This problem does not arise for the parametric learning model discussed earlier since all the available information about (the constant) \( \theta \) is used efficiently in each period.
\[ m^t(z_0) = \sum_{j \in L} \left[ w_j^t m^t(z_{j-1}) + w_j^t y_{j-1} \right] + w^t m^t(z_0) \]

\[ \sum_{j \in L} (w_j^t + w_j^y) + w^t = 1, \quad w_j^t > 0, \quad w_j^y > 0, \quad \forall j, \quad w^t > 0 \]

where, \( m^t(z_0) \) = peasant’s best estimate of the expected yield in period, \( t \).
\( m^t(z_{j-1}) \) = estimate of neighbor \( j \)’s expected yield, obtained from his previous-period acreage allocation decision. Recall here that the parametric specification of the acreage function is assumed to be common knowledge, permitting an estimate of the expected yield to be derived from each neighbor’s acreage decision.
\( y_{j-1} \) = neighbor’s previous-period yield realization. Recall from the specification of the yield function in equation (1) that this will provide an unbiased estimate of the neighbor’s expected yield, \( m(z_{j-1}) \).
\( m^t(z_0) \) = exogenous signal which is assumed to provide an unbiased estimate of the peasant’s current expected yield, \( m(z_0) \).
\( w_j^t, w_j^y, w^t \) = learning weights placed on expected yield estimates obtained from neighbors’ acreage decisions, neighbors’ yield realizations, and the exogenous signal, respectively.

The learning weights in equation (6) correspond to the kernel function in nonparametric estimation. They are, in general, computed by minimizing a loss function that includes both variance and bias.\(^{48}\) As before, neighbors are ranked by distance from the peasant, in type space, to derive the implications of the nonparametric learning model. The learning weight placed on a neighbor depends on the bias associated with the use of his expected yield estimate in the peasant’s learning function, as well as on the quality of the neighbor’s signal. As discussed above, the bias associated with the use of neighbors’ expected yield estimates is increasing in their distance from the peasant in type space. Distant neighbors thus have less influence in the learning function. In addition, following the discussion on parametric learning, signal quality is assumed to be uncorrelated with a neighbor’s distance (in type

\( I \) will test for the presence of multi-period learning in the estimation procedure.

\(^{48}\) See Härdle, 1990, for a discussion on kernel estimation. Manski, 1993a, also uses nonparametric estimation to motivate decision-making under similar conditions.
space) from the peasant. Learning weights in the unrestricted nonparametric learning model are consequently monotonically declining in neighbor rank.

**Learning Model 3**: Nonparametric estimation with a monotonicity restriction (linear smoothing)

As discussed previously, the pattern of estimated learning weights is inconsistent with the implications of both canonical learning models derived above. Consistent implications are, however, obtained with a learning model that places an apparently reasonable, monotonicity, restriction on the yield function, $m'(z_o) > 0$. The peasant’s objective is now to determine the favorability of his location in type space, relative to his neighbors, as well as the suitability of his geographic location to the new technology. To accomplish this, the peasant constructs a (local) linear relationship between expected yield and type, with information obtained from his neighbors. This relationship permits him to estimate the expected yield corresponding to his particular type and geographic location.

The monotonicity restriction can be interpreted as a local linear relationship between expected yield and type, since a linear function is a first-order approximation to any arbitrary function. To derive the implications of this interpretation, we begin our discussion with a stronger assumption, that the expected yield is in fact linear in type. To simplify the exposition, ignore information from neighbors’ previous-period yield realizations and the exogenous signal. Furthermore, assume that all neighbors provide signals of equal quality.\(^{50}\)

---

\(^{49}\) For an estimate obtained from a neighbor’s acreage decision, the quality of the signal depends on the precision and bias in his own expected yield estimate, $m'(z_{o_t})$. For an estimate obtained from a neighbor’s yield realization, the quality of the signal depends on his yield precision, $\tau_{j_t}$. Note here that the yield realization provides an unbiased estimate of the neighbor’s expected yield, $m(z_{o_t})$. While the quality of a neighbor’s signal may be increasing in his type, it is unlikely to be correlated with distance (in type space) from the peasant. This follows since low and high types are equally likely to be distant neighbors in the learning function.

\(^{50}\) Recall, from our previous discussion, that the quality of the signal obtained from a neighbor’s acreage decision is determined by the precision and bias associated with the neighbor’s estimate of his own expected yield. Similarly, the quality of a neighbor’s yield signal depends on the yield precision. The assumption that all neighbors provide signals of equal quality is clearly restrictive; one would expect high types to provide superior information. However, in general, high and low types are equally likely to be distant neighbors, implying that signal quality will be uncorrelated with rank. This assumption is consequently unlikely to affect the implications of the learning model that are finally derived.
peasant’s type to zero, \( z_a = 0 \), the peasant will obtain his expected yield, \( m^A(z_a) \), as the intercept term in an OLS regression of neighbors’ expected yields, \( m^A(z_{j-1}) \), on type, \( z_{j-1} \).

\[
m^A(z_a) = Em^A(z_{j-1}) - \alpha Ez_{j-1}
\]

\[
\alpha = \frac{\frac{1}{n} \sum_{j \in L^a} (z_{j-1} - Ez_{j-1}) m^A(z_{j-1})}{\frac{1}{n} \sum_{j \in L^a} (z_{j-1} - Ez_{j-1})^2}
\]

\[
Em^A(z_{j-1}) = \frac{1}{n} \sum_{j \in L^a} m^A(z_{j-1}) \quad Ez_{j-1} = \frac{1}{n} \sum_{j \in L^a} z_{j-1}
\]

The learning function derived in equation (7) is contrasted with the unrestricted nonparametric learning function derived earlier in equation (6). Equation (6) implicitly assumes \( \alpha = 0 \) since the monotonicity restriction was ignored. In this case, \( \alpha \) is the slope of the regression line. Equation (7) can be simplified to express the peasant’s expected yield estimate, \( m^A(z_a) \), as a linear function of neighbors’ expected yield estimates, \( m^A(z_{j-1}) \).

\[
m^A(z_a) = \sum_{j \in L^a} w^A_j m^A(z_{j-1})
\]

\[
w^A_j = \frac{1}{n} \left[ 1 - \frac{Ez_{j-1} + (z_{j-1} - Ez_{j-1})}{\frac{1}{n} \sum_{j \in L^a} (z_{j-1} - Ez_{j-1})^2} \right]
\]

Introducing information from neighbors’ yield realizations and the exogenous signal, equation (8) has the same structure as the unrestricted nonparametric learning function, equation (6).

The alternative nonparametric learning functions have the same structure since the peasant’s objective in both cases is to determine his expected yield, \( m(z_a) \). It is also easy to verify that learning weights in both models sum to one. The alternative models differ, however, in the pattern of learning weights placed on neighbors’ signals. Unrestricted nonparametric learning
implies positive learning weights, monotonically decreasing in neighbor rank. As I will show below, linear smoothing implies (possibly) negative learning weights and nonmonotonicity.

The pattern of learning weights implied by the linear model can be derived from equation (8). There are two cases to consider; \( z_{it} < \text{EZ}_{jt-1} \) (Figure 1(a)) and \( z_{it} > \text{EZ}_{jt-1} \) (Figure 1(b)).

For \( z_{it} < \text{EZ}_{jt-1} \): \( w^\wedge_j > 0 \) if \( z_{jt} < \text{EZ}_{jt-1} \) and \( w^\wedge_j > 0 \) or \( < 0 \) if \( z_{jt} > \text{EZ}_{jt-1} \)

For \( z_{it} > \text{EZ}_{jt-1} \): \( w^\wedge_j > 0 \) or \( < 0 \) if \( z_{jt} < \text{EZ}_{jt-1} \) and \( w^\wedge_j > 0 \) if \( z_{jt} > \text{EZ}_{jt-1} \)

Consequently, inspection of Figure 1 reveals that \( z_{it} \) lies in the region with positive learning weights, \( w^\wedge_j > 0 \), in both cases. The general implication is that near neighbors will have positive weights, while the sign of the learning weight on intermediate and distant neighbors is ambiguous. Negative weights are more likely to be obtained as a neighbor's distance from the peasant, in type space, increases.

A simple intuition for the implication derived above can be obtained from Figure 1(a).

Assume, for now, that the peasant has the lowest type among his neighbors, \( z_{it} < z_{jt-1}, \forall j \).

The Y-axis consequently lies to the left of all the data points in Figure 1(a). Recall that these data points represent neighbors' expected yields. The peasant obtains his expected yield as the intersection of a regression line, through the data points in Figure 1(a), with the Y-axis.

Consider first the effect of an increase in the most distant neighbor's expected yield on the point of intersection of the regression line. There are two effects here; the regression line shifts up and counter-clockwise. The first effect is obtained because an increase in a neighbor's expected yield is always a positive signal regarding the suitability of the particular geographic location to the new crop. Recall that all neighbors are assumed to share the same locational characteristics. This effect appears as the first right-hand-side term in equation (7).

The second effect is observed because an increase in a distant neighbor's expected yield, conditional on near neighbors' expected yields, is a negative signal about the favorability of the peasant's position in type space, relative to his neighbors. An increase in the slope of the regression line implies an increase in the sensitivity of expected yield to type. Since the peasant has the lowest type among his neighbors, this will cause a downward revision in his expected yield estimate. This effect appears as the second term on the right-hand-side of equation (7), distinguishing the linear smoothing model from the unrestricted nonparametric
model presented earlier. The two effects described above work in opposite directions and the net effect on the point of intersection of the regression line is consequently ambiguous. This result implies that learning weights on distant neighbors have ambiguous sign.

Consider next the effect of an increase in the closest neighbor's expected yield. In this case the regression line will shift up and rotate clockwise. It is easy to verify, following the argument outlined above, that the point of intersection of the regression line will now shift upward. Learning weights on near neighbors are unambiguously positive. The implications derived from the graphical exercise are consequently consistent with those derived, more generally, earlier.

It is easy to verify, from equation (8), that learning weights are monotonically declining in type for \( z_q < Ez_{j-1} \), and monotonically increasing in type for \( z_q > Ez_{j-1} \). The implications of the linear smoothing model must, however, be derived in terms of neighbor-rank, rather than by type, to be consistent with the implications of the canonical learning models, obtained earlier. There are two special cases to consider. For the first special case, where the peasant has the lowest (highest) type among his neighbors, it is easy to verify from Figure 1 that a monotonically declining pattern of learning weights in neighbor-rank is obtained, with the possibility of negative weights on intermediate and distant neighbors. This follows because neighbor-ranks are monotonically increasing (decreasing) in type for this extreme case. For the second special case, with \( z_q = Ez_{j-1} = 0 \), all neighbors have equal and positive weight in the learning function, from equation (8). This follows because the second term on the right hand side of equation (7) is now absent since \( Ez_{j-1} = 0 \). For the more general case, with an intermediate peasant and \( z_q \neq Ez_{j-1} \), the pattern of learning weights is bounded by the implications of the special cases described above. Learning weights will in general be declining in neighbor-rank, although nonmonotonicity in the range of positive weights (on near neighbors) may be obtained in some instances.\(^{51}\)

\(^{51}\) To see this, consider the case in which the peasant's type, \( z_q \), is closest to the second-lowest neighbor's type in Figure 1(a). Recall that learning weights are monotonically decreasing in type, in this case (with \( z_q < Ez_{j-1} \)). Transforming the implication, in terms of neighbor-rank, the learning weight on the closest (in type space) neighbor will consequently be smaller than the weight on the second-ranked neighbor, who has the highest type and hence the highest weight, in this case. The pattern of monotonically declining learning weights on intermediate and distant neighbors, \( z_{j-1} > Ez_{j-1} \), will nevertheless continue to be obtained, since neighbor-ranks are increasing in type for
In estimating the pattern of learning weights across neighbors, the econometrician does not control for the peasant's position in type space, relative to his neighbors. The estimated pattern of learning weights is consequently a (matrix) weighted average of the alternative patterns described above. Nevertheless, a monotonically declining pattern of learning weights in neighbor-rank, with the possibility of negative weights on intermediate and distant neighbors, is quite generally obtained.

While the assumption that neighbors’ expected yields are linear in type provides a useful bench-mark case, a more flexible local linearity assumption is consistent with the monotonicity restriction. The implications derived above must be adjusted under the local linearity assumption to place less weight on distant neighbors. This adjustment is consistent with the pattern of weights implied by Fan’s (1992) local linear regression smoother. Negative weights on distant neighbors continue to remain a possibility, but the minimum (negative) weight may now be obtained on an intermediate neighbor, with learning weights increasing thereafter.

This discussion concludes the description of the alternative learning models proposed in the paper. To summarize the implications derived from these models, learning weights are positive and monotonically declining in neighbor rank with unrestricted nonparametric learning, whereas all neighbors have equal influence with parametric learning. The linear smoothing model implies positive weights on near neighbors and (possibly) negative weights on intermediate and distant neighbors. Distant neighbors now have less influence in the learning function due to the local linearity assumption. The patterns of learning, weights corresponding to the alternative learning models are displayed in Figure 2.

3. The econometrician’s inference problem

The econometrician’s objective is to infer from the peasant’s observed acreage decisions, and the prior acreage decisions and yield realizations of his neighbors, how the peasant himself

such neighbors. The range of nonmonotonicity increases as the peasant’s type, z_i, shifts closer to E_{z_{i+1}}. However, this effect is weakened by the increasing tendency to approximate a pattern of equal learning weights across neighbors as z_i tends towards E_{z_{i+1}}. Recall that for the second special case described above, z_i = E_{z_{i+1}} = 0, all neighbors have equal influence in the peasant’s learning function.

84
inferred. Recall from our previous discussion that the peasant's objective with nonparametric learning is to determine his expected yield, \( m(z_u) \), directly. In contrast, with parametric learning the peasant's objective is to learn the value of the locational parameter, \( \theta \). The learning functions derived earlier were consequently expressed in terms of \( m(z_u) \) or \( \theta \), depending on the learning model under consideration. The econometrician's first objective is to nest the implications of the alternative learning models to permit direct testing of these models against each other.

I now assume that the yield function is linear and that the acreage function is additively separable in expected yield.\(^{52}\) Furthermore, I assume that the econometrician obtains the best available estimates of the parameters of the acreage and yield functions, assuming of course that these functions are correctly specified by her. This follows since the econometrician has access to considerably more information than the peasant did historically. The peasant updated his assessment of the new technology in each period with information received from a limited sample of neighbors. The econometrician, at least in principle, has access to a panel data-set containing the entire history of acreage allocations and yield realizations, of all the cross-sectional units in the system. To simplify the exposition, I will assume below that the econometrician's estimates correspond to the true values of the acreage and yield function parameters.

To begin with, consider the yield function specified earlier in equation (1). Following the assumptions listed above, equation (1) is re-written as,

\[
y_u = m(z_u) + \eta_u = \theta + \beta z_u + \eta_u
\]  

(1')

Following equation (1'), \( m(z_u) = \theta + \beta z_u \), for the discussion below. The econometrician's estimated yield function may be expressed as,

\[
y_{\hat{u}} = \theta + \beta z_{\hat{u}} + \epsilon_{\hat{u}}
\]  

(1'')

where, \( \epsilon_{\hat{u}} \) = estimated residual from the econometrician's yield function.

\(^{52}\) There is little to be gained by maintaining additional generality at this stage since linear yield and acreage functions are specified, in any case, in the econometrician's estimation procedure. I will, however, later discuss how possible misspecification of these functions is controlled for in the estimation procedure.
Recall here that the econometrician’s estimates are assumed to correspond to the true values of the yield function parameters.

The peasant’s best estimate of his expected yield, \( m(z_u) \), with nonparametric learning is \( y_u \). Similarly, his best estimate of \( \theta \) with parametric learning is \( \theta_u = y_u - \beta z_u \). Recall that \( \beta \) is assumed to be known to the peasant. From equation (1’), the econometrician’s yield residual, \( \epsilon'_u \), can then be interpreted as peasant "error." This error is interpreted as the deviation in the peasant’s historical estimate, from \( m(z_u) \) or \( \theta \) depending on the learning model that was utilized by him. From equation (1’’) we obtain,

\[
\epsilon'_u = y_u - m(z_u) = \theta'_u - \theta
\]  

Turning next to the acreage function and following the assumptions listed above, equation (2) can be re-written as,

\[
A_u = m^A(z_u) + h(.) = \theta'_u + \beta z_u + h(.)
\]  

where, \( h(.) \) subsumes all the remaining terms in the acreage function in equation (2).

The econometrician’s estimated acreage function may be expressed as,

\[
A_u = \theta + \beta z_u + h(.) \ast \epsilon'_u
\]  

The residual obtained from the econometrician’s estimated acreage function can be interpreted, as before, as the deviation from \( m(z_u) \) or \( \theta \), depending on the learning model under consideration. The peasant’s "error" in this case is a consequence of the limited information that was available to him and reflects his state of learning at a given point in time. Comparing equation (2’) and equation (2’’),

\[
\epsilon_u^A = m^A(z_u) - m(z_u) = \theta_u - \theta
\]  

Inspection of equation (i) and equation (ii) suggests an econometric strategy that nests the implications of the alternative learning models, derived earlier. Transforming the learning functions in terms of deviations, from \( m(z_u) \) and \( \theta \) respectively, the econometrician’s estimated acreage and yield residuals can be used to test the alternative learning models.
directly. To see this, begin with the transformation of the parametric learning function.

Noting that learning weights sum to one, equation (5) is re-written as,

\[ (\theta^A_n - \theta) = \sum_{j \in L_n} \left( w_j^A (\theta^A_{j-1} - \theta) + w_j^Y (\theta^Y_{j-1} - \theta) \right) + w^Y (\theta^Y_n - \theta) \]  

(5’)

Substituting from equation (i) and equation (ii),

\[ \epsilon^A_n = \sum_{j \in L_n} \left( w_j^A \epsilon^A_{j-1} + w_j^Y \epsilon^Y_{j-1} \right) + u_n \]  

(9)

The nonparametric learning function may be transformed in a similar manner. Recall here that the learning functions for both unrestricted nonparametric estimation and linear smoothing display the same structure. Noting as before that learning weights sum to one, equation (6) is re-written as,

\[ m^A(z_{uj}) - m(z_{uj}) = \sum_{j \in L_n} \left[ w_j^A \left[ m^A(z_{uj}) - m(z_{uj}) \right] + w_j^Y \left[ y_{j-1} - m(z_{uj}) \right] \right] + w^Y \left[ m^Y(z_{uj}) - m(z_{uj}) \right] \]  

(6’)

Substituting from equation (i) and equation (ii),

\[ \epsilon^A_n = \sum_{j \in L_n} \left( w_j^A \epsilon^A_{j-1} + m(z_{uj}) - m(z_{uj}) \right) + w_j^Y \left( \epsilon^Y_{j-1} + m(z_{uj}) - m(z_{uj}) \right) + u_n \]  

(9’)

Using the econometrician’s estimated acreage and yield residuals to test the implications of the nonparametric model leads to measurement error in this case since the \([m(z_{uj}) - m(z_{uj})] \) terms on the right hand side of equation (9’) are omitted. The variance of this measurement error is increasing in distance (in type space), since \([m(z_{uj}) - m(z_{uj})] \) is larger in absolute magnitude for distant neighbors. The attenuation bias associated with measurement error is consequently increasing in distance.\(^{53}\) Since both nonparametric learning models imply less weight on distant neighbors, the omission of the \([m(z_{uj})-m(z_{uj})] \) term simply accentuates this implication. The measurement error can consequently be ignored in this case.

\(^{53}\) This implication requires that the true regressors \([m^A(z_{uj}) - m(z_{uj})] \) and \([y_{j-1} - m(z_{uj})] \) be orthogonal across neighbors.
Ignoring the measurement error associated with nonparametric estimation, the learning weights are not affected in the transition from the original learning functions to equation (9). Note here that equation (9) and equation (9') are identical if the \([m(z_{t-1}) - m(z_0)]\) terms are ignored. The implications of the alternative learning models derived previously, and presented in Figure 2, consequently remain unchanged. The modified learning function, equation (9), nests the implications of the alternative learning models, permitting direct testing of the underlying learning process.

The econometric procedure is implemented in two stages. In the first stage, acreage and yield are estimated as functions of exogenous characteristics such as irrigated area, rainfall, input and output prices, and asset ownership. Recall here that district-level data are obtained over a 22 year period for the econometric exercise. While the models discussed earlier considered learning at the individual level, the district is treated as the decision-making unit in the econometric procedure. Issues associated with potential aggregation bias will consequently be addressed later in the paper. The first stage estimation procedure essentially controls for variation in exogenous characteristics, which would independently lead to spatial diffusion of the new technology over time. Conditioning for changes in these exogenous characteristics, the first-stage residuals are interpreted as deviations, from \(m(z_0)\) or \(\theta\), as discussed above. Changes in the spatial pattern of residuals over time reveal the underlying structure of the learning process. To test the alternative models of learning, the district's acreage residual is regressed on its geographically contiguous neighbors' previous-period acreage and yield residuals, following equation (9).\(^{54}\) Ranking neighbors by their distance from the district in type space, the pattern of coefficients on neighbors' residuals reveals the underlying learning process. Learning weights are estimated in this case by assuming that the same learning model is utilized by all districts, over the entire time period under consideration.

---

\(^{54}\) The district's learning set is defined as the set of geographically contiguous neighbors in this paper. The basic idea here is that contiguous districts are likely to display the same geographical attributes, such as soil quality. However, this definition ignores information access issues associated with, for instance, the quality of transportation links connecting districts. Geographically contiguous neighbors may, in some cases, provide very little information to a district. One solution to control for such access issues is to only consider neighbors that face similar input and output prices in the learning set. Such neighbors are likely to be included in the same geographic market as the district, which implies in turn that transportation links are well developed.
The econometric procedure described above gives rise to an alternative interpretation for the estimated pattern of learning weights.\textsuperscript{55} Recall that the estimated coefficients on neighbors’ residuals in the second-stage revealed a pattern with positive weights on near neighbors and negative weights on intermediate and distant neighbors. Attenuation of learning weights on distant neighbors was also observed. The estimated pattern of learning weights was shown to be consistent with the linear smoothing model in the previous section. As I will discuss below, this pattern is also consistent with parametric learning, when both $\theta$ and $\beta$ are unknown.

Relaxing the assumption that $\beta$ is known, equation (2') with parametric learning can be written as,

\[ A_n = \theta_n + \beta_n z_n + h(\cdot) \]  

\text{(2'')}

Comparing equation (2'') and equation (2'''),

\[ \epsilon_n' = (\theta_n + \beta_n z_n) - (\theta_* z_n) \]  

\text{(iii)}

The acreage residual is now interpreted as the deviation of the peasant’s expected yield estimate, $m^*(z_n)$, from the true expected yield, $m(z_n)$. A similar exercise interprets the yield residual as the deviation from $m(z_n)$. This interpretation of the acreage and yield residuals may be contrasted with the previous interpretation of these residuals, with parametric learning. Recall, from equation (i) and equation (ii), that the residuals were previously interpreted as deviations from $\theta$.

With the new interpretation of the first-stage residuals, the econometrician can no longer use equation (5'), in which the learning function is expressed in terms of $\theta$, to test for parametric learning. The appropriate learning function is now expressed in terms of $m(z_n)$, corresponding to equation (6'). While the implication of equal learning weights across neighbors is no longer obtained, parametric learning continues to be distinguished from nonparametric learning since $m^*(z_n)$ is obtained from an underlying parametric learning process (which is unobserved by

\text{\footnotesize \textsuperscript{55} I am grateful to Kevin Lang for bringing this interpretation to my attention.}}

89
the econometrician). This contrasts with nonparametric learning in which the peasant directly estimates his expected yield.

The pattern of learning weights associated with the parametric model in equation (6') may be directly obtained from the previous discussion on linear smoothing. Note that linear smoothing is essentially local parametric learning which assumes linearity in the yield function but allows $\theta$ and $\beta$ to be unknown. The type elasticity, $\beta$, corresponds to the slope coefficient, $\alpha$, in the linear smoothing model. The pattern of learning weights that we are interested in is consequently directly obtained from equation (8). Recall, from the previous discussion, that this pattern is characterized by positive weights on near neighbors and (possibly) negative weights on intermediate and distant neighbors. Learning weights are also monotonically declining in neighbor-rank. This last implication distinguishes the parametric learning model from linear smoothing. Linear smoothing assumes local linearity, which implies less weight on distant (in type space) neighbors.

Parametric learning may nevertheless be consistent with the estimated pattern of learning weights since attenuation on distant neighbors follows, for a number of reasons, from the two-stage econometric procedure. For instance, the measurement error associated with the transition from equation (6') and equation (9'), to equation (9), leads to increased attenuation in the learning weights on higher ranked neighbors. Furthermore, bias associated with the second-stage estimation procedure, discussed in the next section, will also result in such attenuation. In general, it may not be possible to distinguish between linear smoothing and parametric learning, when $\theta$ and $\beta$ are unknown, with the two-stage econometric procedure. The principle empirical result of the paper, that peasants appear to condition for differences in their neighbors characteristics, nevertheless remains unaffected by the alternative interpretation of the estimated pattern of learning weights.

4. Estimation results

District-level data from 270 districts are obtained over a 22 year period, 1966-87.\textsuperscript{56} The data

\textsuperscript{56} Districts are fairly large administrative units, both in terms of area as well as population, comprising hundreds of villages. District lines typically coincide with historical administrative boundaries and it is usually possible to characterize a district, as a unit, in terms of its wealth, entrepreneurial talent, and propensity to adopt new ideas.
are obtained from official Government of India sources. These include Directorate of Economics and Statistics publications, census reports, Indian Agricultural Statistics, and Ministry of Shipping and Transport publications. Evenson, Pray, and Rosegrant (1994) provides a detailed description of the data and its construction.

As discussed earlier, the estimation procedure is implemented in two stages. In the first stage, the acreage allocated to the new HYV technology and the yield are estimated as functions of the district's exogenous characteristics. In the second stage, residuals from the estimated acreage and yield functions are used to test the implications of the alternative learning models.

4.1 First-stage estimation
Determinants of acreage and yield in the first-stage include extension services, roads, literacy, asset ownership (bullock and tractor density), wages, output prices, irrigated area and rainfall. Means and standard deviations of these variables are computed over four time periods, 1967-72, 1972-77, 1977-82, 1982-87, in Table 1. All the variables increase over these time periods, almost without exception. This observation suggests an alternative explanation for the diffusion of HYV technology that follows from the increased availability of agricultural inputs and relatively favorable output prices. As discussed earlier, the first-stage regressions control for this effect, permitting the alternative learning models to be tested in the second stage.

Turning to the acreage variables, the percentage of Gross Cropped Area (GCA) allocated to the HYV crops increases substantially over time. GCA here refers to the total area cultivated over all cropping seasons. Although rice is cultivated far more extensively than wheat in India when considering acreage allocated to both traditional varieties and HYV, the proportion of GCA allocated to HYV is roughly the same for the two crops. This result is consistent with the general impression that the diffusion of HYV technology has proceeded more rapidly with wheat than with rice.
While the aggregate statistics suggest an increase in the acreage allocated to HYV crops, the spatial distribution of the new technology is important for our purpose. A simple graphical exercise permits us to (crudely) observe the spatial diffusion of the new technology over time. Districts that allocate more than a threshold share of the GCA to the HYV crop (0.05) are marked with a cross, +, in Figures 3(a) and 3(b). + sizes are increasing in the share of GCA allocated to the new technology in the following ranges: 0.05-0.1, 0.1-0.2, 0.2-0.3, 0.3-0.5, 0.5-1.0. It is apparent from Figures 3(a) and 3(b) that subsequent to the Green Revolution in the 1960s, HYV technology has spread over most of the wheat and rice growing areas of the country. It has been adopted as an entirely new crop in many instances as well. Apart from this spatial diffusion, the intensity of allocation of the new crop also appears to have increased, judging from the darkening of the scatter-plots in Figures 3(a) and 3(b) over time. The extensive spread of the new technology observed in Figures 3(a) and 3(b) is reassuring since the tests of the alternative learning models would have little power otherwise.

Following the HYV acreage trend, HYV wheat and rice yields, in tons/hectare, have also increased over the sample period. The slight decline in rice yield in the 1982-87 period is probably due to the low rainfall in those years. The observed increase in yield over time is somewhat surprising since one would expect areas with less favorable attributes to adopt the new technology, at the margin, within a district. This result suggests that improvements in the HYV technology, or the factor input increase that we discussed above, dominate the adverse effect of the new adopters' inferior attributes.

The included determinants of the acreage and yield functions are fairly standard and appear in most agricultural supply-response studies (see for instance Bapna,Binswanger, Quizon, 1981). The social learning interpretation for technology diffusion is also broadly consistent with the Nerlovian adjustment model, in which the peasant reaches his optimal acreage with a

---

57 Prices of variable inputs, bullock and tractor density, irrigated area and rainfall, are included in logs. Shift variables, such as extension, literacy and roads, appear in levels. This follows the standard treatment of technological progress variables in the estimation of production functions. The gross irrigated area and the literacy variable are normalized by GCA. Finally, I include yield in logs as the dependent variable in the yield function, but specify (acreage/GCA) in levels as the corresponding variable in the acreage function. This permits the inclusion of observations (districts) with zero acreage allocated to HYV in a given year.
lag. In this case the lag is a consequence of the peasant's lack of information about the new technology.

The first-stage regressions are estimated with district and year fixed-effects. The data are first-differenced to sweep-out district fixed-effects, while the year fixed-effects are retained as dummies to be estimated. First-differencing is preferred to within-estimation in this case because of the presence of lagged output prices in the first-stage equations, consistent with the Nerlovian adaptive expectations model (see Nowshirvani, 1968, for a discussion). Although output price and acreage/yield are jointly determined, the endogeneity problem is avoided with lagged output prices, if the residuals are serially uncorrelated. In our case, the residuals are highly serially correlated. First-differencing effectively pre-whitens the residuals, eliminating the endogeneity problem. Level residuals, required for the second-stage estimation, are ultimately recovered from the differenced residuals obtained from the first-stage regressions.

---

58 See Nerlove, 1958, for a discussion. Behrman, 1966, and Nowshirvani, 1968, are examples of early supply response studies that incorporate this approach. The first-stage acreage regression does not explicitly include lagged acreage, which is a standard term in the Nerlovian model. This term would, however, appear if the learning function was estimated by substituting expressions for the residuals, from the first stage.

59 Consider the simple model,

\[ y_t = \alpha P_{t-1} + \epsilon_t \]

\[ E(P_{t,1}) \neq 0 \quad \epsilon_t = \epsilon_{t-1} + u_t \]

where \( u_t \) is white-noise.

Substituting for \( \epsilon_t \), it is easy to see that the orthogonality condition fails, resulting in inconsistent \( \alpha \) estimates. First-differencing we obtain,

\[ (y_t - y_{t-1}) = \alpha (P_{t-1} - P_{t-2}) + u_t \]

Consistent estimates of \( \alpha \) are now obtained since,

\[ E(\epsilon_{t-1} | u_t) = E(\epsilon_{t-2} | u_t) = 0 \]

60 It is straightforward to obtain level residuals from first-differenced residuals in this case. Consider the model,

\[ y_{it} = z_{it} \beta + g_t \gamma + f_i \delta + \epsilon_{it} \]

where, \( g_t \) = year dummies

\( f_i \) = district fixed-effects.

First-differencing we obtain,

93
Tables 2 and 3 present OLS and 2SLS estimates of the first-stage regressions. One would generally expect wages and acreage/yield to be jointly determined, resulting in an endogeneity problem. Following Benjamin (1992), I include rural population and aggregate state wages as instruments for the district wage. The first-stage coefficients may be treated as nuisance parameters since we are ultimately interested in estimates of the second-stage learning weights. I will consequently discuss the results of the first-stage estimation procedure very briefly below.

The first point to note from Tables 2 and 3 is that own prices (first and second lags) play a relatively unimportant role in the acreage allocation decision. This result is somewhat surprising, given the strong price-effects typically obtained in supply-response studies (see, for instance, Behrman, 1966). The price effect does, however, increase in the yield function, especially for rice. Prices of competing/complementary crops are also included in the acreage and yield functions, with mostly insignificant effects.\textsuperscript{61} Finally, 2SLS estimates of the wage coefficient are significantly negative in both the acreage and the yield regressions, with the

\[(y_{u_{i}}-y_{u_{i-1}})=\xi z_{u_{i}}+\xi g_{i}+\xi \varepsilon_{u_{i-1}}\]

Estimates of \(\beta\), \((g_{i}, g_{i})\), \((\varepsilon_{u_{i}}, \varepsilon_{u_{i}})\), are obtained from the first-stage regressions. Estimates of \(g\) are derived by setting \(g_{0}=0\). We can consequently compute a consistent estimate of \(f_{i}\), since the time-series is fairly long (22 years), using these first-stage estimates.

\[\hat{f}_{i}=y_{i}-z_{i}\beta-g\]
\[y_{i}=\frac{\sum y_{u_{i}}}{T}, \quad z_{i}=\frac{\sum z_{u_{i}}}{T}, \quad g=\frac{\sum g_{i}}{T}\]

Consistent estimates of the level residuals can then be derived from the expression,

\[\hat{\varepsilon}_{u_{i}}=y_{u_{i}}-z_{u_{i}}\beta-g_{i}\hat{f}_{i}\]

These computed level residuals are used in the second-stage regressions.

\textsuperscript{61} The idea here is that the peasant must distribute his resources among various crops, in the presence of input and total acreage constraints. An increase in the price of a competing crop would lead to a shift, both in acreage as well as in variable inputs, to that crop. An increase in the price of a complementary crop would relax the liquidity constraint, increasing the resources available to the peasant. Maize and sugarcane prices are included in the rice regressions, while gram prices are included in the wheat regressions. An increase in gram price has a significantly negative impact on wheat acreage, while an increase in maize price reduces rice yields. None of the other price effects are significant.
exception of wheat yield. The other important input price, that of fertilizer, is ignored since it does not vary across districts in a given year. Fertilizer prices are consequently subsumed in the year fixed-effects in the acreage and yield regressions.

The second point to note in Tables 2 and 2 is that extension services have a significant impact on acreage allocation. The corresponding (weakly) negative coefficient in the yield function may be a consequence of intra-district heterogeneity. Extension services presumably induce peasants with less favorable attributes to adopt the new technology, leading to a decline in average yield.

The third point to note is that irrigation and rainfall play an unambiguously positive and significant role in both the acreage allocation decision as well as the yield function. Rainfall is more important for rice yields, whereas irrigation is more important for wheat yields. Both variables are important determinants of acreage for rice and wheat.

The Box-Pearson Q statistic for serial correlation and the Moran I statistic for spatial correlation are also included in Tables 2 and 3. We can easily reject the null hypotheses of

\[ Q = \frac{\sum \hat{r}_{t-1}}{\sum \hat{\epsilon}_{t-1}^2} \]

The advantage of the Q statistic over the more commonly used Durbin-Watson statistic is that it provides an unbiased test for the presence of serial correlation even with lagged dependent variables as regressors. As discussed earlier, the second-stage regression includes the district's previous-period acreage residual as a regressor. The Q statistic allows us to consistently test for serial correlation across both estimation stages. Furthermore, the first-stage

---

62 HYV cultivation generally tends to be more labor intensive than the cultivation of traditional varieties and alternative crops. The negative effect of wages on HYV acreage may then be explained as the outcome of a shift to competing crops or traditional varieties. 2SLS estimates, not reported here, were also obtained using rural population as the only instrument for wages. Omission of aggregate state wages as an instrument had little effect on the estimation results.

63 This result is consistent with the well documented evidence that extension services play an important role in the adoption of new technology (see for instance, Feder, Lau, and Slade, 1987).

64 Current rainfall is highly positively correlated with acreage allocation. One explanation for this, somewhat curious, observation is that peasants usually wait for the first few showers before sowing the crop. This initial period may be sufficient for them to (correctly) predict the rainfall in a given cropping season.

65 The Q statistic is expressed as,
no serial correlation and no spatial correlation, on the basis of the computed Q and MI statistics.\textsuperscript{66} This is an expected result since these correlations are implied by the second-stage learning model. Since standard errors are now inconsistent, I extend the White and Newey-West standard errors to account for spatial correlation. Standard errors robust to heteroscedasticity, serial correlation, and spatial correlation, are computed in practice by replacing non-zero terms in the error variance-covariance matrix by sample statistics, derived in turn from the estimated residuals (see White, 1981, for a discussion).

4.2 Second-stage estimation

As discussed previously, the district’s acreage residual is regressed on its neighbors’ (and its own) previous-period acreage and yield residuals in the second-stage. These residuals are obtained from the estimated first-stage acreage and yield functions. Geographically contiguous neighbors are ranked by distance in type space in the second stage to test the implications of the alternative learning models.

\begin{equation}
Q \text{ statistic is computed using level residuals since the second-stage regressions are run with residuals in levels.}
\end{equation}

The Moran I statistic is expressed as,

\[
MI = \frac{\sum_i \sum_j \hat{e}_i w_{ij} \hat{e}_j}{\sum_i \hat{e}_i^2/d}
\]

where, \( J \) = total number of (pairs of) contiguous neighbors in the system
\( d \) = total number of districts
\( w_{ij} = 1 \) if districts \( i \) and \( j \) are neighbors, and 0 otherwise.

The MI statistic is asymptotically normally distributed under the null hypothesis of no spatial correlation. The expressions for the first two moments are provided in Case (1991). The MI statistic is computed using first-differenced residuals, which are approximately serially uncorrelated, for each year. The individual MI statistics are then summed to obtain the statistics presented in Tables 2 and 3. The corresponding moments of the normal distribution under the null are computed accordingly.

\textsuperscript{66} The relatively large MI statistic for the yield residuals in Table 3 may be due to the simplifications we employ to construct the district yield series. Yield data are only available at the state level. The underlying district yields are derived from the state level data, implying that yields within a state will be highly correlated. Most of a district’s neighbors tend to be from the same state which would lead in turn to highly spatially correlated yield residuals. The construction of the district yield series will be discussed in detail later in the paper.
Recall from the previous discussion on the identification problem that social learning can be distinguished from unobserved correlated effects since it occurs with a single-period lag in our agricultural application. Neighbors’ contemporaneous acreage residuals are nevertheless included in the second-stage to control for these correlated effects, as well as to proxy for omitted variables in the first stage.\(^{67}\) Separating the district’s own previous-period acreage and yield residuals from its neighbors’ residuals, equation (9) can be written as,

\[ \epsilon_{it}^{\prime} = w_0 \epsilon_{i,t-1}^{\prime} + w_0' \epsilon_{i,t-1}^{\prime} + \sum_{j \in T, j \neq i} (\rho_j \epsilon_{j,t}^{\prime} + w_j \epsilon_{j,t-1}^{\prime} + w_j' \epsilon_{j,t-1}^{\prime}) + u_{it} \]  

(10)

where, \( w_0, w_0' \) = learning weights on district’s own previous-period acreage and yield residuals, respectively.
\( \rho_j \) = coefficient on neighbor’s current acreage residual.

Preliminary regression results, not reported here, revealed significant serial correlation in the residuals of equation (10). Serial correlation and spatial correlation lead to inconsistent estimates in this case since the lagged dependent variable and neighbors’ dependent variables are included as regressors. A general solution to this problem is to introduce additional lags to remove serial correlation and, similarly, to include neighbors’ neighbors residuals to purge (first-order) spatial correlation. I begin by quasi-differencing equation (10) until serial

---

\(^{67}\) In addition to controlling for unobserved correlated effects, neighbors current residuals also proxy for omitted variables in the first-stage estimation procedure. For instance, they may partially control for the forward-looking component of the acreage allocation decision, associated with experimentation. Following Foster and Rosenzweig, 1994, with strategic experimentation the peasant’s decision is determined by his own, and his neighbors’, current assets; conditional on the previous experience with the new technology. The asset terms capture the forward-looking component of the acreage decision here. In my case, assets are correlated with type, by definition. Since high types are likely to receive superior information, and consequently to learn faster, the first-stage residuals will in turn be correlated with type. Recall that these residuals are interpreted as reflecting the state of learning. The current residuals in the second-stage estimation consequently proxy for the asset terms that would enter as determinants of acreage in a forward-looking model.
correlation is eliminated. In this case the elimination of serial correlation by the inclusion of additional lags also implies that spatial correlation in the residuals is absent.\textsuperscript{68}

Equation (10) required the inclusion of two additional lags before serially uncorrelated errors were obtained. One interpretation of this result is that \( u_n \) in equation (10) follows an AR2 process.\textsuperscript{69}

\[
\begin{align*}
    u_n &= \delta_1 u_{n-1} + \delta_2 u_{n-2} + \nu_n \\
\end{align*}
\]

The structural form of the quasi-differenced second stage regression equation is obtained from equations (10) and (11) as,

\begin{itemize}
\item \textsuperscript{68} The basic result here is that quasi-differencing will not remove serial correlation when spatial correlation is also present. To see this consider the simple model,

\[
\begin{align*}
    e''_n &= w_0 e''_{n-1} + \rho_j \epsilon_j + w_j e''_{j-1} + u_n \\
    u_n &= \delta_1 u_{n-1} + \delta_2 u_{n-2} + \nu_n \\
\end{align*}
\]

where district, \( i \), has a single neighbor, \( j \), \( \nu_n \) is white noise, and yield realizations are ignored as sources of information. Quasi-differencing we obtain,

\[
\begin{align*}
    e''_n &= (w_0 + \delta_1) e''_{n-1} + \rho_j \epsilon_j + (w_j - \delta_1) \epsilon_j - \delta_1 w_j + (w_j - \delta_1) \epsilon_j + \delta_2 u_{n-2} + \nu_n \\
\end{align*}
\]

Since \( u_n \) is also serially correlated, the transformed residual continues to be serially correlated. The ability to remove serial correlation with quasi-differencing consequently implies the absence of spatial correlation.

\item \textsuperscript{69} Recall, from equation (9) and equation (9'), that \( u_n \) is interpreted as the deviation from \( m(z_n) \) or \( \theta \), for the exogenous signal. If we think of this signal as being provided by the local extension worker, then it is quite likely that the deviation will decrease over time as he too learns about the new technology. An alternative interpretation for the two additional lags is that learning is characterized by a multi-period rule; either because underlying disaggregate learning itself follows a multi-period rule (which could be the case with nonparametric learning) or because aggregation problems result in a multi-period aggregate learning even when disaggregate learning follows a single-period rule. While it is not possible to test for this alternative interpretation, the assumption that the extension worker’s state of learning is serially correlated does appear reasonable.

98
\[ \epsilon_i = (w_0 \delta_1 - \delta_1 w_0) \epsilon_{i-1} + \delta_2 \epsilon_{i-2} \epsilon_{i-3} + w_1 \epsilon_{i-1} \epsilon_{i-2} - \delta_2 \epsilon_{i-2} \epsilon_{i-3} \]

\[ \sum_{j \in I, j \neq i} \left[ \rho \epsilon_j + (w_j - \delta_1 \rho) \epsilon_{j-1} - (\delta_1 w_j + \delta_2 \rho) \epsilon_{j-2} - \delta_2 w_j \epsilon_{j-2} \epsilon_{j-3} + w_j \epsilon_{j-1} \epsilon_{j-2} - \delta_2 w_j \epsilon_{j-2} \epsilon_{j-3} \right] \epsilon_{i-1} \epsilon_{i-2} \epsilon_{i-3} \]

Since equation (12) is linear in variables, the reduced form specification is estimated with OLS. The structural coefficients, which we are interested in, are estimated with Nonlinear Least Squares (NLS).

It is convenient in this case to test the over-identifying restrictions with the Likelihood Ratio statistic since both unrestricted OLS and restricted NLS estimates are easily obtained. The results of the Likelihood Ratio test provide information about the presence of the aggregation problem, associated with the use of district-level data, which was discussed earlier. The basic aggregation result applied here is that aggregate learning appears to follow a multi-period rule even when underlying disaggregate learning is characterized by a single-period rule, when aggregation problems are present (see the Appendix for a discussion). The regression equation derived in equation (12) implicitly ignores aggregation problems. If such problems are present, then learning weights will enter beyond the first lag, contaminating the reduced form coefficients in the second and third lags. The structural specification will consequently fail the overidentifying restrictions.

It is indeed likely that aggregation problems are present and, consequently, that the overidentifying restrictions will be rejected by the data. Such an outcome does not necessarily imply that the estimation results obtained with aggregate data are completely undermined. Confidence in the econometric results would depend on the severity of the aggregation bias. I consequently compare learning weights obtained with restricted and unrestricted NLS estimation. Inspection of equation (12) reveals that the structural coefficients are just identified when coefficients on neighbors' second-lagged and third-lagged residuals \( (\epsilon_{j-2}, \epsilon_{j-3}, \epsilon_{j-2}, \epsilon_{j-3}, j \in R_i, j \neq i) \) are estimated in their reduced-form. If the model were correctly specified,

---

70 As discussed earlier, multi-period disaggregate learning rules can be obtained with the nonparametric model. Such multi-period learning would also lead to contamination in the reduced form coefficients on second and third lags, since a single-period learning rule is implicitly assumed in deriving equation (12).
without aggregation problems, then the unrestricted NLS estimates of the structural coefficients in equation (12), with coefficients on neighbors’ second-lagged and third-lagged residuals estimated in reduced form, would be consistent with restricted estimates obtained with all three lags. The difference between the restricted and unrestricted structural estimates would increase with the severity of the aggregation bias. I consequently estimate structural coefficients in equation (12), with and without neighbors’ extended lags. Estimation results presented below reveal similar learning weights for the two cases, although the over-identifying restrictions are formally rejected, suggesting that while aggregation problems may be present they do not seriously undermine the estimation results.

The results of the second-stage estimation with OLS, 2SLS, and adjusted 2SLS, residuals from the first-stage are presented for rice and wheat in Tables 4(a) and 4(b). The adjusted 2SLS estimates correct for censoring in the first-stage acreage regression using a procedure suggested by Greene (1981).\textsuperscript{71} As discussed previously, the alternative models of learning presented earlier place restrictions on the pattern of learning weights by neighbor rank (distance in type space). To test these implications, each district’s contiguous neighbors are ranked by distance, in each year, in equation (12).\textsuperscript{72}

\textsuperscript{71} The OLS and 2SLS estimates presented in Table 2 do not correct for the censoring associated with observed zero HYV acreage, except to exclude districts that allocate no HYV over the entire sample period. Such potential outliers are omitted since the censoring problem is presumably most acute, with the latent variable, A\textsubscript{it}, far below zero, for districts that never adopt the new technology. With regard to more formal solutions to the censoring problem, Honoré (1992) presents a semiparametric estimator to obtain consistent estimates of censored regressions with fixed-effects. Honoré’s estimator is consistent when error terms are jointly normal, with equal variance (for a given individual over time) but arbitrary positive correlation. Following the previous discussion, the residuals from the first-stage acreage regressions are interpreted as deviations, (θ\textsuperscript{th}-θ) or [m\textsuperscript{th}(z\textsubscript{it})-m(z\textsubscript{it})], which reflect the state of learning. The variance of θ\textsuperscript{h}\textsubscript{it} or [m\textsuperscript{h}(z\textsubscript{it})-m(z\textsubscript{it})] decreases over time as the peasant learns about the new technology (for parametric learning, for instance, this is apparent from equation (4)). Without homoskedasticity over time (for a given individual), inconsistent estimates will be obtained. I consequently implement an alternative approach, suggested by Greene (1981), in which estimated OLS coefficients are divided by the proportion of nonlimit (non-zero) observations in the sample to correct for censoring. This correction is an exact result when the regressors are multinormally distributed. This condition is clearly not satisfied in our case because of the fixed-effects dummies. Greene’s Monte Carlo simulation results suggest, however, that the correction is fairly robust to non-normality in the regressors, including the presence of dummy variables.

\textsuperscript{72} The distance metric is computed as the square of the difference between the district’s type in the current year and the neighbor’s type in the previous year. A district’s type in a given year is, in turn, constructed as a weighted sum of its exogenous characteristics. These characteristics are readily identified in this case as the determinants of acreage and yield in the first stage. With regard to the weights to be applied, the estimated coefficients of the acreage and yield function appear as natural candidates. Since it is not clear that the peasant parametrizes the yield
The closest seven neighbors are included in each district's learning function since very few districts in our sample have eight or more contiguous neighbors. For districts with less than seven contiguous neighbors zeros are inserted for missing neighbors, both for acreage as well as yield residuals. The yield function is estimated with district-years in which strictly positive acreage was allocated to the HYV crop. Zeros are similarly inserted for the yield residual in other district-years. Zeros are interpreted in both cases described above as providing no information about the new technology. Since missing neighbors do not belong in the learning function, no weight should be placed on such observations when they are included. The estimation procedure, however, treats all observations (including zeros) symmetrically, assigning the same weight to all observations in a given rank. This procedure leads to attenuation in the estimated learning weights for neighbor ranks which display a significant number of zeros. Missing neighbors are assigned high ranks in the second-stage estimation procedure. Estimated learning weights on higher ranks will consequently be biased towards zero. This attenuation is of little consequence when the underlying learning structure is nonparametric, since neighbors' learning weights are monotonically declining in rank (distance). However, the econometrician can no longer distinguish between linear smoothing and parametric learning, when $\theta$ and $\beta$ are unknown, in this case.\textsuperscript{73}

\textsuperscript{73} Although the results are not reported here, I also performed robustness tests to ensure that the results of the second-stage estimation procedure were not driven by choice of the number of neighbors (seven) or by the inclusion of zeros for missing observations. I estimated the model with five neighbors instead of seven, both with zeros for missing observations as well as by removing all districts that had less than five contiguous neighbors. The estimation results in both cases were qualitatively similar to those obtained with seven neighbors and missing observations, which are reported later in Tables 4(a) and 4(b).

An alternative strategy that avoids problems with an unbalanced panel would be to derive implications of the alternative learning models in terms of a neighbor's influence, as a function of his distance from the peasant in type space rather than his rank. For instance, learning weights on a neighbor could be expressed as a flexible (polynomial) function of his distance from the peasant. With such a specification, the coefficients of the distance function rather than the rank coefficients would be estimated in the second stage, combining terms across neighbors. This would avoid problems associated with an unbalanced panel, but leave the estimation results sensitive to the choice of distance metric.
As discussed earlier, the structural coefficients in equation (12) are estimated with NLS. The first seven rows in Tables 4(a) and 4(b) list coefficients, $\rho_j$, corresponding to neighbors’ current residuals, $\epsilon^A_{p,j}$, $j = 1, \ldots, 7$, which were included in equation (10), and consequently in equation (12), to control for contemporaneous correlated effects and omitted variables in the first-stage. The next eight rows contain learning weights, $w^A_{0}, w^A_{j}$, associated with neighbors’ (and the district’s own) previous-period acreage residuals, $\epsilon^A_{p-1}, \epsilon^A_{p-1}$, $j = 1, \ldots, 7$. Since we are especially interested in the pattern of learning weights, this section of coefficients is presented in bold-face type in Tables 4(a) and 4(b). Finally, the coefficient, $w^y_{0}$, associated with the district’s own previous-period yield residual, $\epsilon^y_{p-1}$, is also included in these Tables, together with estimates of the serial correlation coefficients, $\delta_1$ and $\delta_2$, which appear in equation (12). The coefficients on the remaining yield residuals are mostly insignificant and do not reveal any systematic pattern. These coefficients are not reported in Tables 4(a) and 4(b) to preserve space. I will provide an explanation for the weak influence of the yield residuals below.

Most of the coefficient estimates in Tables 4(a) and 4(b) are statistically significant. Coefficients on neighbors’ current acreage residuals are, without exception, positive and highly statistically significant. The estimated coefficients are monotonically declining in neighbor rank (distance). Since no structure is placed on the contemporaneous correlated effects, I do not attempt to interpret this result. Learning weights, associated with neighbors’ (and the district’s own) previous-period acreage residuals, are initially monotonically declining in neighbor rank (distance). Beyond some intermediate rank, which displays negative weight, the pattern reverses and learning weights increase thereafter, although weights on distant neighbors continue to remain negative. As discussed earlier, this nonmonotonic pattern, with positive weights on near neighbors and negative weights thereafter, broadly coincides with the implications of the nonparametric linear smoothing model, as well as of the parametric learning model with unknown $\theta$ and $\beta$. This constitutes the principal empirical result of the paper. The same pattern of learning weights is obtained, without exception, for all the alternative specifications presented and for both crops. Recall here that residuals from OLS, 2SLS, and adjusted 2SLS, estimates from the first stage are used in the second stage. Furthermore, most of the estimated learning weights are statistically significant. As discussed previously, restricted and unrestricted NLS estimates will differ widely when aggregation
problems are severe.\textsuperscript{74} Inspection of restricted and unrestricted estimates in Tables 4(a) and 4(b), however, suggests that the results are not seriously undermined by aggregation bias.\textsuperscript{75}

Finally, the estimated coefficient on the district’s previous-period yield residual is positive and statistically significant in all cases. As mentioned above, however, the coefficients on the remaining yield residuals do not reveal any particular pattern in neighbor rank and are statistically insignificant. One explanation for this result is that yield realizations provide relatively imprecise information about the new technology and consequently have less influence in the learning function. Acreage residuals contain information that is indirectly received from a wide range of districts over the entire history of acreage allocations and yield realizations. In contrast, yield residuals contain information from a single yield realization.\textsuperscript{76} In addition, only state-level yield data for HYV rice and wheat are available. District-level data are constructed by assuming that the ratio of HYV yield to traditional-variety yield, computed with state level data, holds over all districts within that state in a given year. This assumption is unlikely to be satisfied in practice. Inconsistent estimates will be obtained in the first-stage yield regression if the HYV to traditional-variety yield ratio is a function of the district’s type.\textsuperscript{77} The weak yield effect, however, allows us to ignore the potential inconsistency associated with the construction of the yield data.

\textsuperscript{74} This result is also obtained if disaggregate learning is characterized by a multi-period rule, since we implicitly assume single-period learning in deriving equation (12).

\textsuperscript{75} I was unable to obtain convergence with unrestricted NLS estimation for wheat. The second-lag was consequently estimated in structural form, to obtain partially-restricted estimates in Table 4(b) rather than (strictly) unrestricted estimates.

\textsuperscript{76} To clarify these statements, consider the acreage precision term, \( \tau^A_{it} \), in equation (4). The expression for \( \tau^A_{it} \) can be derived recursively forward, for each period from \( t = 0 \), to obtain,

\[
\tau^A_{it} = \sum_{r=0}^{t} \tau^S_{fr} + \sum_{r=0}^{t-1} \sum_{j \neq i} \sum_{r=0}^{t-1} (1-r) (\tau^j_r + \tau^j_{fr})
\]

This expression is derived by ignoring neighbors’ neighbors’ signal and yield precisions. It nevertheless includes the entire history of neighbors’ (and the peasant’s own) signal and yield precisions. Contrast the expression for the acreage precision, \( \tau^A_{it} \), above, with the yield precision, \( \tau^Y_{it} \).

\textsuperscript{77} This follows because the error in construction (difference between the true yield and the constructed yield), which will then be a function of the district’s type, enters the error term in the first-stage yield regression. The orthogonality conditions are consequently rejected since yield is specified as a function of the district’s type.
Standard errors in Tables 4(a) and 4(b) do not require correction to account for the use of estimated first-stage residuals in the second-stage regressions. Newey and McFadden (forthcoming) show in general that second-stage standard errors require correction if and only if consistency of the first stage estimator affects consistency of the second-stage estimator. It is straightforward to show that second-stage standard errors require no correction under fairly weak conditions in our case. This result may also explain why the estimated second-stage coefficients are so robust to the alternative first-stage specifications that we experimented with.

While the pattern of estimated learning weights is broadly consistent with the implications of the linear smoothing model, as well as of the parametric model with unknown \( \theta \) and \( \beta \), inference about the specific structure of the learning process should be treated with caution. The negative coefficients on intermediate and distant neighbors’ signals do, however, suggest that the peasant conditions, albeit crudely, for differences in his neighbors’ types when learning from them. Inference about the underlying learning process depends, of course, on the validity of the identifying restrictions placed on the learning function. Recall that neighbors’ current residuals were included in the second-stage estimation procedure to proxy

\[ y_{it} = z_{it} \beta + \epsilon_{it} \quad \text{First-Stage} \]

\[ \epsilon_{it} = \rho \epsilon_{it-1} + u_{it} \quad \text{Second-Stage} \]

The moment conditions are,

\[
\frac{1}{T} \sum_{t} z_{it} \epsilon_{it} = \frac{1}{T} \sum_{t} z_{it} (y_{it} - z_{it} \beta) = 0
\]

\[
\frac{1}{T} \sum_{t} \epsilon_{it-1} u_{it} = \frac{1}{T} \sum_{t} (y_{it-1} - z_{it-1} \beta) [y_{it} - z_{it} \beta] - \rho [y_{it-1} - z_{it-1} \beta] = 0
\]

We need to show that the second-stage moment condition is unaffected by first-stage inconsistency. Specifically, it is easy to verify that,

\[
\frac{1}{T} \sum_{t} \frac{\partial}{\partial \beta} (\epsilon_{it-1} u_{it}) = 0
\]

when the gradient is evaluated at the true \( \beta \), implying that the first-stage moment condition is satisfied, and assuming \( E(z_{it-1} u_{it}) = E(z_{it} \epsilon_{it-1}) = E(z_{it} \epsilon_{it}) = 0 \). The last equality is directly implied by the first-stage moment condition. In addition, the district’s type, \( z_{it} \), is assumed to be exogenous in the paper, implying that it is uncorrelated with unobserved information shocks, \( u_{it} \) and \( \epsilon_{it-1} \).
for unobserved correlated effects, as well as for omitted variables, in the first-stage regressions. Consistency of the estimated learning weights is consequently contingent on the econometrician's ability to control for these correlated effects and omitted variables in the second-stage regression.

The identifying restrictions are rejected when the omitted variables are both spatially and serially correlated. This occurs, for instance, when inputs such as credit, seeds, or fertilizer, are rationed and their availability increases over time. In this case, high types with their superior resources are likely to gain access to these scarce inputs before low types. Treating access to scarce inputs as an unobserved shock, low types' residuals could be correlated with high types' lagged (previous-period) residuals, even if learning were absent. A similar correlation is obtained when the rate of learning-by-doing is (positively) correlated with type and the state of learning is unobserved. Such interpretations of sequential adoption are, however, characterized by an asymmetry, in the sense that late adopters' residuals are correlated with early adopters' previous-period residuals. In contrast, the learning model described in this paper does not restrict attention to the process by which late adopters learn from early adopters. All districts are treated symmetrically in the second-stage estimation procedure, regardless of when they adopt the new technology, weakening the asymmetric effect described above.\textsuperscript{79}

Failure of the identifying restrictions may also follow from aggregation bias associated with the contemporaneous effects. Following our previous discussion, correlated effects will spill over into subsequent periods (lags) when aggregation problems are present. The correlation between a district's current residual and its neighbors' previous-period residuals may simply be an outcome of this aggregation spill-over. Aggregation problems associated with the learning weights were tested for earlier by comparing restricted and unrestricted NLS estimates in equation (12). A similar exercise can be performed in this case by comparing unrestricted OLS and restricted NLS estimates of $\rho_j$ in equation (12). Similar estimates are obtained in this case suggesting that the aggregation spill-over may not be significant.

\textsuperscript{79} Although the results are not reported here, I tested for the presence of the asymmetric effect by splitting the sample of districts into low and high types (using predicted yield as the index for type). The estimation results were generally unaffected by this partitioning of the data.
7. Conclusion

Technology diffusion is interpreted in this paper as the outcome of a social learning process in which agents learn from their neighbors about the quality of a new and uncertain technology. Agents condition their decisions on their neighbors' previous decisions and experiences, since these provide information about the new technology. While the literature is consistent with regard to the motivation for social learning, there is little agreement on the structure of the learning process. Determining this structure is ultimately an empirical problem. In this case we are faced with a double inference problem; to infer from the agent's observed actions, and the prior actions and outcomes of those around him, the manner by which the agent himself infers. My strategy in this study is to choose a convenient application, derive alternative learning models, and then determine which of these models is most consistent with the observed pattern of technology diffusion. I choose the adoption of Green Revolution technology in Indian agriculture for the empirical exercise.

Alternative models of social learning may be broadly distinguished by the peasant's ability to extract information from neighbors whose characteristics (types) differ from his own. When learning is efficient, the peasant is able to condition for differences in neighbors' types to learn about the new technology. Consequently, all neighbors have equal influence in the peasant's learning function, regardless of their distance from him in type space. The conditioning described above uses knowledge of the production process which, if incorrect, would result in biased inference about the quality of the new technology. Consequently, the agent may rationally prefer to use relatively inefficient learning rules that discount valuable information from distant (in type space) neighbors, but place little structure on the production function. The implications of the alternative learning models are distinguished by the effect of distance (in type space) on a neighbor's influence in the learning function.

The implications of the alternative learning models are tested with district-level data over a 22 year period. The estimation results are consistent with fairly efficient learning behavior in which the peasant conditions, albeit crudely, for differences in neighbors' types to learn about the new technology. This constitutes the principal empirical result of the paper. This result suggests that peasant behavior may not be entirely norm-driven, at least with regard to investment decisions. In addition, it implies that learning models that ignore heterogeneity,
and treat neighbors as a group, may be misspecified since learning weights vary across neighbors (by distance in type space).

Optimal policy, aimed at spreading the new technology, depends crucially on the structure of the learning process. Our empirical result suggests that learning is fairly efficient, implying that information will endogenously diffuse quite rapidly through the system, even when population groups are heterogeneous. This result implies a role for policy, in increasing interaction among such groups. While social learning occurs spontaneously in agriculture, since investment decisions and outcomes are readily observable, it is difficult to identify institutions that facilitate social learning in other industries. Indian dairy cooperatives appear to have played a role in promoting social learning by increasing interaction among producers, but in a competitive environment where there is no private cost associated with the dissemination of information (Munshi and Parikh, 1994). For most other industries, the challenge for policy lies in inducing rival firms to share information.\footnote{Compensating early adopters is an obvious solution in this case (see Feder and Slade, 1985, for a discussion). An alternative strategy would be to move investment behavior toward the cooperative equilibrium by creating institutions which allow firms to internalize the information externality.}
References


Appendix: Aggregating learning rules

The learning models described in the paper assume learning at the individual level. The implications of the alternative learning models are, however, tested with aggregate, district-level, data. This section describes the aggregation problems that are potentially associated with such an econometric exercise. Solutions to this problem are then derived. I finally present the main aggregation result of the paper; aggregate learning will, in general, appear to follow a multi-period rule even when disaggregate learning is characterized by a single-period rule. I begin this section with a stylized, 3-peasant 2-district, example in which learning is assumed to occur at the farm level. District 1 consists of peasants 1 and 2, while peasant 3 is the sole member of district 2. Period-1 learning in this model is specified as,

\[ \epsilon_{11} = w_{11} \epsilon_{10} + w_{12} \epsilon_{20} + w_{13} \epsilon_{30} + z_{11} \]

\[ \epsilon_{21} = w_{21} \epsilon_{10} + w_{22} \epsilon_{20} + w_{23} \epsilon_{30} + z_{21} \]

\[ \epsilon_{31} = w_{31} \epsilon_{10} + w_{32} \epsilon_{20} + w_{33} \epsilon_{30} + z_{31} \]  

(1)

where, \( \epsilon_{it} \) = peasant i’s acreage residual in period, t (deviation from \( m(z_o) \) or \( \theta \)).
\( z_o \) = information shock received by peasant, i, in period, t. In the context of our model this shock includes yield residuals, \( \epsilon_{i,t-1} \), and exogenous information, \( u_{it} \).

w’s are learning weights.
The true aggregate relationship is,

\[ (\epsilon_{11} + \epsilon_{21}) = (w_{11} + w_{21}) \epsilon_{10} + (w_{12} + w_{22}) \epsilon_{20} + (w_{13} + w_{23}) \epsilon_{30} + (z_{11} + z_{21}) \]  

(2a)

\[ \epsilon_{31} = w_{31} \epsilon_{10} + w_{32} \epsilon_{20} + w_{33} \epsilon_{30} + z_{31} \]  

(2b)

The aggregate relationship to be estimated, since only district-level decisions are observed, is,

\[ (\epsilon_{11} + \epsilon_{21}) = k_{11} (\epsilon_{10} + \epsilon_{20}) + k_{12} \epsilon_{30} + (z_{11} + z_{21}) \]  

(3a)

\[ \epsilon_{31} = k_{21} (\epsilon_{10} + \epsilon_{20}) + k_{22} \epsilon_{30} + z_{31} \]  

(3b)

Comparing equations (2a) and (3a), equations (2b) and (3b),
\[ k_{11} = \frac{\epsilon_{10}(w_{11} + w_{21}) + \epsilon_{20}(w_{12} + w_{22})}{\epsilon_{10} + \epsilon_{20}} \quad k_{12} = (w_{13} + w_{23}) \]  

\[ k_{21} = \frac{\epsilon_{10}w_{31} + \epsilon_{20}w_{32}}{\epsilon_{10} + \epsilon_{20}} \quad k_{22} = w_{33} \]  

(4a)  

(4b)

The aggregate learning weights are obtained as a weighted average of the underlying disaggregate learning weights. The aggregate learning weights will, in general, vary over time even when the disaggregate learning weights are time-stationary since they are functions of the \( \epsilon \)'s. The practical difficulty with this situation is that we will be unable to estimate the coefficients on neighbors' residuals with aggregate time-series data. In addition, aggregation bias is also associated with this problem, in the context of our spatial learning model. As the new technology diffuses slowly across a large area, aggregate data will tend to dilute the learning effect, at least initially.

To see this consider a special case, of the stylized single-period example that we presented earlier, in which the 3 peasants are now located along a line and peasants learn from contiguous neighbors only. Suppose now that in period-0 only peasant 1 receives a positive signal about the new technology (\( \epsilon_{10} > 0 \), \( \epsilon_{20} = \epsilon_{30} = 0 \)). Assume further that peasant 3 places positive weight on peasant 2 but zero weight on peasant 1 (\( w_{32} > 0 \), \( w_{31} = 0 \)). From equation (4b) we will consequently obtain \( k_{21} = 0 \). This incorrectly implies that district 2 does not learn from district 1 and, therefore, that social learning is absent. In fact we would have obtained \( k_{21} > 0 \) if we had specified instead, \( \epsilon_{20} > 0 \).

The dilution effect occurs, in general, when peasants in the learning district observe only a limited set of peasants in the adjacent target district who, presumably, are located close to the common border of the two districts. In this situation, if the new technology appears initially in the unobserved area, then (aggregate) acreage in the target district will increase without a corresponding acreage increase in the learning district. Notice, however, that the new technology will ultimately reach the learning district, albeit with a time-lag. Aggregate learning coefficients would thus increase over time as the new technology diffused through the system. I will later use this intuition to propose a solution to the aggregation problem, as well as to test for its presence.

The previous discussion underscores the importance of deriving time-stationary aggregate learning coefficients. Two general approaches are proposed to achieve this time-stationarity. The first approach follows from equations (4a) and (4b) and places restrictions either on the disaggregate learning weights or on the distribution of farm-level residuals. The second approach follows from our discussion on the dilution effect, specifying multi-period aggregate learning even when disaggregate learning follows a single-period rule.

The conditions on disaggregate residuals are easier to derive. For \( \epsilon_{10} = \epsilon_{20} \) we obtain, The intra-district homogeneity condition requires that all peasants in the target district have the same residual. This condition is unlikely to be satisfied in practice since each peasant has a unique learning set. Information about the new technology will consequently vary across
\[ k_{11} = \frac{w_{11} + w_{21} + w_{12} + w_{22}}{2} \quad k_{12} = (w_{13} + w_{23}) \quad (4a') \]

\[ k_{21} = \frac{w_{31} + w_{32}}{2} \quad k_{22} = w_{33} \quad (4b') \]

peasants implying, in turn, different residuals.

The second set of conditions places restrictions on the disaggregate learning weights. For \( w_{11} + w_{21} = w_{12} + w_{22}, \ w_{31} = w_{32}, \)

\[ k_{11} = w_{11} + w_{21} = w_{12} + w_{22} \quad k_{12} = w_{13} + w_{23} \quad (4a'') \]

\[ k_{21} = w_{31} = w_{32} \quad k_{22} = w_{33} \quad (4b'') \]

As I will show below, the general condition that we require is that the sum of the weights placed on any two peasants in a district by the peasants in any district (including their own) must be the same.

The conditions on learning weights, or on the distribution of residuals, necessary to obtain single-period aggregate learning are quite stringent. In most instances, all the information necessary to characterize disaggregate learning cannot be obtained in a single period at the aggregate level. In such cases, this additional information can be obtained by including multiple lags at the aggregate level, even when disaggregate learning is characterized by a single-period rule. The underlying disaggregate learning rule is progressively revealed over multiple periods and, in principle, it is always possible to fully characterize disaggregate learning if data over a sufficiently long time-period are available. The intuition for this solution is that information about the new technology ultimately is revealed at the district level even if it is not immediately apparent. While aggregate learning rules may not contain sufficient information to fully characterize important intra-district heterogeneity in the target district, this heterogeneity is ultimately revealed with multi-period rules. Aggregate learning consequently appears to follow a multi-period rule even with single-period disaggregate learning.

To clarify the preceding discussion we return to our stylized example with three peasants and two districts. We equate the right-hand-sides of equations (2a) and (3a), comparing coefficients of \( \epsilon_{10}, \ \epsilon_{20}, \ \epsilon_{30}, \) to obtain 3 equations with 2 unknowns, \( k_{11} \) and \( k_{12}. \) A similar exercise with equations (2b) and (3b) also yields 3 equations with 2 unknowns, \( k_{21} \) and \( k_{22}. \) We are thus unable to solve uniquely for the aggregate coefficients, as functions of the disaggregate learning weights. However, unfolding aggregate learning rules can be used to characterize the unobserved underlying disaggregate learning rules in this case. Introducing multiple lags in the aggregate learning rule increases the number of unknowns, allowing us to
derive a unique solution for the aggregate coefficients. For our simple example, true aggregate period-2 learning is expressed as,

\[(e_{12} + e_{22}) = (w_{11} + w_{21})e_{11} + (w_{12} + w_{22})e_{21} + (w_{13} + w_{23})e_{31} + (z_{12} + z_{22}) \]  
\[
(2a')
\]

\[e_{32} = w_{31}e_{11} + w_{32}e_{21} + w_{33}e_{31} + z_{32} \]  
\[
(2b')
\]

With a multi-period learning rule the aggregate relationship to be estimated is,

\[(e_{12} + e_{22}) = k_{11}(e_{11} + e_{21}) + k_{12}e_{31} + k_{13}(e_{10} + e_{20}) + (z_{12} + z_{22}) \]  
\[
(3a')
\]

\[e_{32} = k_{21}(e_{11} + e_{21}) + k_{22}e_{31} + k_{23}(e_{10} + e_{20}) + z_{32} \]  
\[
(3b')
\]

Substituting for \(e_{11}, e_{21}, e_{31}\) from equation (1), and comparing right-hand-sides as before, we obtain 3 equations in 3 unknowns in each case, allowing us to completely characterize disaggregate learning. The multi-period aggregate learning rules are now time-stationary since aggregate coefficients are functions of disaggregate learning weights only.

It is easy to verify that the shocks, \(z's\), do not cancel out when comparing right-hand-sides of equations (2a') and (3a'). The same result is obtained when equations (2b') and (3b') are compared. The potential problem here is that the underlying disaggregate learning will not be revealed with multi-period aggregate learning when shocks enter the system in each period. In our case, however, shocks entering the system are assumed to be orthogonal to the acreage residuals, \(\epsilon's\). Since the model is estimated over a long time-period, significantly greater than the number of lags in the learning function, we will continue to obtain consistent, time-stationary, estimates of the aggregate coefficients.

I conclude this section with a more general description of the aggregation problem and the main aggregation result.

There are \(n\) peasants aggregated in \(m\) districts, \(m < n\).
Let \(\epsilon_0 \in \mathbb{R}^n\) be period-0 residuals at the farm level.
Let \(\epsilon_1 \in \mathbb{R}^n\) be the corresponding period-1 residuals.
Period-1 residuals are conditioned on period-0 residuals.
The learning rule is characterized by the linear transformation, \(L: \mathbb{R}^n \rightarrow \mathbb{R}^n, \epsilon_1 = L\epsilon_0 + z_1\).
Here the elements of \(L\) are the learning weights placed on neighbors (and the peasant himself),
\(z_1 \in \mathbb{R}^n\) is a vector of shocks, orthogonal to \(\epsilon_0\). As discussed above, \(z_1\) includes yield residuals and the exogenous signal.

I also characterize the general aggregation rule in which farm-level residuals are summed to derive district-level residuals, \(A: \mathbb{R}^n \rightarrow \mathbb{R}^m\).
Let \(E_0 \in \mathbb{R}^n\), \(E_1 \in \mathbb{R}^n\) be the aggregate residuals at the district level.
Then, \(E_0 = Ae_0, E_1 = Ae_1\).
I now derive conditions on disaggregate learning rules such that a time-stationary, single-period, aggregate learning rule, \( K: \mathbb{R}^m \to \mathbb{R}^m \), is obtained relating \( E_t \) and \( E_0 \). This learning rule is specified as,

\[ E_t = K E_0 + Z_t \quad (5) \]

\( Z_t \in \mathbb{R}^m \) is a vector of aggregate shocks, \( Z_t = A z_t \)

Substituting for \( E_t \) and \( E_0 \) in equation (1) we obtain,

\[ A L e_0 = K A e_0 \quad (6a) \]

Time-stationary learning implies rules independent of \( e_0 \) and, hence, must satisfy the condition,

\[ A L = K A \quad (6b) \]

\( A L \) reflects the information that is required to fully characterize disaggregate learning. \( K A \) reflects the information that is actually available from aggregate learning behavior. Let \( N_t \) be the number of peasants in district, \( s, s = 1, \ldots, m \). It is easy to verify that \( K A \) is an \( m \times n \) matrix in which the first \( N_t \) columns are identical, the next \( N_2 \) columns are identical, and so on. We will refer to this pattern, henceforth, as the "block-column" structure. In general, however, \( A L \) does not display any particular pattern. It would thus be impossible to solve for the \( m^2 \) elements of \( K \) by comparing corresponding elements in \( A L \) and \( K A \) unless \( A L \) also exhibited a block-column structure. In that case there would be \( m \) distinct columns in both matrices, implying \( m^2 \) equations and \( m^2 \) unknowns and, hence, a unique solution for \( K \). Our discussion on matrix structure is easily interpreted in terms of the information contained in \( A L \) and \( K A \). The block-structure observed in \( K A \) is a consequence of the limited information available from aggregate learning behavior since all peasants are weighted equally at the district level. There is thus insufficient information to fully characterize disaggregate learning unless we place restrictions on \( A L \) as well.

To derive the necessary restrictions on disaggregate learning we partition \( L \) into sub-matrices with \( N_t \) rows and \( N_c \) columns, \( c, r = 1, \ldots, m \). As before, \( N_t \) refers to the number of peasants in the first district, and so on. It is easy to verify the necessary and sufficient condition for the structure described above,

\[ \sum_{i=1}^{N_t} w_{ij} = w_{rc} \quad \forall j = 1, \ldots, N_c, \quad c, r = 1, \ldots, m \quad (7) \]

where, \( w_{ij} \) is the weight placed on peasant \( j \) by peasant \( i \).

This restriction implies that the sum of the elements in each column of each sub-matrix of \( L \) must be equal. The condition must be independently satisfied in each sub-matrix. The interpretation of this restriction in the context of farm-level learning is that the sum of the weights placed on any two peasants in a district by the peasants in any district (including their own) must be the same. When this condition is satisfied, the aggregate learning rule takes the simple form,

\[ K = A L A' (A A')^{-1} \]
It is easy to verify that \((AA')^{-1}\) is a diagonal matrix with \(1/N_s\), \(s=1,\ldots,m\), as the diagonal terms.

When the conditions on learning weights, or on the distribution of residuals, necessary to obtain single-period learning fail to be satisfied, time-stationary learning weights may still be obtained by including multiple lags at the aggregate level. As discussed earlier, aggregate learning will then appear to follow a multi-period rule even with single-period disaggregate learning. In terms of the AL, KA comparison that we discussed earlier, conditioning current decisions on multiple past decisions generalizes the KA term, increasing the information available at the aggregate level. Instead of a single K matrix, we now have T such matrices, where T is the number of time-periods included in the aggregate learning rule. The aggregation problem associated with the single-period aggregate learning rule is most severe when all n columns in AL are distinct. In this case there are m*n distinct elements in AL compared with \(m^2\) distinct elements in KA. As discussed earlier, it is then impossible to solve for the \(m^2\) elements of the K matrix that appear in KA, by comparing corresponding elements of KA and AL. By including T=n/m periods in the aggregate learning rule we have \(m^*m^*T=m*n\) elements in the T, K-matrices. Disaggregate learning is then fully characterized.

The disaggregate learning and aggregation rules now take the form,
\[\epsilon_t = L\epsilon_0 + z_1, \quad \epsilon_2 = L\epsilon_1 + z_2, \ldots, \quad \epsilon_T = L\epsilon_{T-1} + z_T\]

By repeated substitution,
\[\epsilon_t = L^t\epsilon_0 + \sum_{r=1}^{t} L^{t-r}z_r\]

E₀ = Aε₀, \(E_t = A\epsilon_t, \ldots, \ E_T = A\epsilon_T\)
Z₁ = Az₁, ..., Zₜ = Azₜ

The multi-period aggregate learning rule is specified as,
\[E_T = \sum_{r=0}^{T-1} K_rE_r + Z_T\]

This can be simplified as,
\[ALTE₀ = \sum_{r=0}^{T-1} K_rAL^{-r}ε₀ + V_T\]  \hspace{1cm} (9a)

\[V_T = \sum_{r=0}^{T-1} K_rA\sum_{t=1}^{r} L^{t-r}z_r + A_{T} - A\sum_{r=1}^{T} L^{T-r}z_t\]

Ignoring \(V_T\) for now, time-stationary learning rules must satisfy the condition,
\[ AL^T = \sum_{t=0}^{T-1} K_t AL^t \] (9b)

AL\(^T\) has m*n distinct elements by definition. \( \Sigma K_t AL^t \) also has m*n distinct elements.\(^{81}\) The unique solution for the, m*m*T=m*n, elements of \( K_t, t=0, \ldots, T-1, \) can then be derived by comparing elements of \( AL^T \) and \( \Sigma K_t AL^t \). Time-stationary aggregate learning coefficients are consequently obtained.

In practice, we will estimate the learning coefficients with a time-series of length greater than T. Since \( V_T \) is orthogonal to \( \epsilon_0 \), by definition, reintroducing \( V_T \) in the learning function does not affect our result. We continue to obtain consistent, time-stationary, estimates of the aggregate coefficients.

---

\(^{81}\) Notice here that \( K_t AL^t, t=1, \ldots, T-1, \) does not display the pattern of identical blocks of columns that we observe with \( K_0 A \), since AL has distinct columns, by definition, in this case.
Figure 1: Pattern in Neighbors’ Expected Yields

(a) zit < Ezjt-1

(b) zit > Ezjt-1
Figure 2: Implications of Alternative Learning Models

learning weight

parametric estimation

unrestricted nonparametric (kernel) estimation

neighbor rank (distance)

nonparametric estimation with a monotonicity restriction (linear smoothing)
Figure 3(a): HYV Rice Share (increasing in + size)
Figure 3(b): HYV Wheat Share (increasing in + size)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HYV Rice/GCA (%)</td>
<td>2.348</td>
<td>5.867</td>
<td>9.367</td>
<td>12.394</td>
</tr>
<tr>
<td></td>
<td>(5.843)</td>
<td>(9.618)</td>
<td>(12.912)</td>
<td>(18.505)</td>
</tr>
<tr>
<td>HYV Wht/GCA (%)</td>
<td>2.301</td>
<td>5.907</td>
<td>9.815</td>
<td>11.729</td>
</tr>
<tr>
<td></td>
<td>(6.085)</td>
<td>(9.618)</td>
<td>(11.955)</td>
<td>(13.664)</td>
</tr>
<tr>
<td>Arice/GCA (%)</td>
<td>22.391</td>
<td>22.816</td>
<td>23.172</td>
<td>24.416</td>
</tr>
<tr>
<td></td>
<td>(23.754)</td>
<td>(23.587)</td>
<td>(23.411)</td>
<td>(26.404)</td>
</tr>
<tr>
<td>Awheat/GCA (%)</td>
<td>11.750</td>
<td>12.957</td>
<td>14.098</td>
<td>15.089</td>
</tr>
<tr>
<td></td>
<td>(12.904)</td>
<td>(13.218)</td>
<td>(14.194)</td>
<td>(15.425)</td>
</tr>
<tr>
<td>Rice Yield</td>
<td>1.132</td>
<td>1.419</td>
<td>1.684</td>
<td>1.660</td>
</tr>
<tr>
<td></td>
<td>(0.887)</td>
<td>(0.992)</td>
<td>(1.440)</td>
<td>(2.936)</td>
</tr>
<tr>
<td>Wheat Yield</td>
<td>1.153</td>
<td>1.362</td>
<td>1.491</td>
<td>1.535</td>
</tr>
<tr>
<td></td>
<td>(0.895)</td>
<td>(0.956)</td>
<td>(1.013)</td>
<td>(1.721)</td>
</tr>
<tr>
<td>Extension/GCA</td>
<td>0.006</td>
<td>0.008</td>
<td>0.017</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.019)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Literacy</td>
<td>0.304</td>
<td>0.337</td>
<td>0.374</td>
<td>0.410</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.096)</td>
<td>(0.102)</td>
<td>(0.109)</td>
</tr>
<tr>
<td>Roads</td>
<td>1621.328</td>
<td>2019.261</td>
<td>2731.759</td>
<td>3275.424</td>
</tr>
<tr>
<td></td>
<td>(1436.269)</td>
<td>(1971.477)</td>
<td>(2999.611)</td>
<td>(3796.747)</td>
</tr>
<tr>
<td>Qbull/GCA</td>
<td>444.442</td>
<td>437.457</td>
<td>406.606</td>
<td>402.472</td>
</tr>
<tr>
<td></td>
<td>(252.355)</td>
<td>(266.971)</td>
<td>(264.903)</td>
<td>(286.103)</td>
</tr>
<tr>
<td>Qtrac/GCA</td>
<td>0.656</td>
<td>1.243</td>
<td>2.277</td>
<td>4.011</td>
</tr>
<tr>
<td></td>
<td>(1.273)</td>
<td>(2.101)</td>
<td>(3.448)</td>
<td>(6.734)</td>
</tr>
<tr>
<td>Wage</td>
<td>2.942</td>
<td>4.518</td>
<td>6.833</td>
<td>11.650</td>
</tr>
<tr>
<td></td>
<td>(1.264)</td>
<td>(2.101)</td>
<td>(3.326)</td>
<td>(5.268)</td>
</tr>
<tr>
<td>Price</td>
<td>96.249</td>
<td>144.050</td>
<td>158.347</td>
<td>190.840</td>
</tr>
<tr>
<td></td>
<td>(37.155)</td>
<td>(58.960)</td>
<td>(61.849)</td>
<td>(63.318)</td>
</tr>
<tr>
<td>Pmaize</td>
<td>65.836</td>
<td>102.691</td>
<td>118.757</td>
<td>158.118</td>
</tr>
<tr>
<td>Psugar</td>
<td>129.682</td>
<td>159.860</td>
<td>207.095</td>
<td>293.410</td>
</tr>
<tr>
<td></td>
<td>(46.985)</td>
<td>(35.797)</td>
<td>(78.433)</td>
<td>(81.722)</td>
</tr>
<tr>
<td>Pwheat</td>
<td>95.479</td>
<td>137.330</td>
<td>163.669</td>
<td>200.916</td>
</tr>
<tr>
<td></td>
<td>(34.914)</td>
<td>(43.195)</td>
<td>(40.066)</td>
<td>(40.249)</td>
</tr>
<tr>
<td>Pgram</td>
<td>100.895</td>
<td>163.642</td>
<td>245.315</td>
<td>376.676</td>
</tr>
<tr>
<td></td>
<td>(25.959)</td>
<td>(35.276)</td>
<td>(56.781)</td>
<td>(90.337)</td>
</tr>
<tr>
<td>GIA/GCA</td>
<td>0.226</td>
<td>0.252</td>
<td>0.290</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>(0.201)</td>
<td>(0.212)</td>
<td>(0.231)</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Rain</td>
<td>1010.359</td>
<td>1026.011</td>
<td>1049.908</td>
<td>1004.926</td>
</tr>
<tr>
<td></td>
<td>(568.068)</td>
<td>(542.380)</td>
<td>(527.639)</td>
<td>(553.397)</td>
</tr>
</tbody>
</table>

NOTE:- Standard deviations are in parentheses.
GCA = Gross Cropped Area (total acreage planted over all cropping seasons)
HYV Rice, HYV Wht = area allocated to HYV rice and wheat, respectively
Arice, Awheat = area allocated to rice and wheat, respectively
Qbull, Qtrac = number of bullocks and tractors, respectively
GIA = Gross Irrigated Area (total area irrigated over all cropping seasons)
Table 2: First-Stage Estimation: Acreage Function

<table>
<thead>
<tr>
<th></th>
<th>Rice OLS</th>
<th>Rice 2SLS</th>
<th>Wheat OLS</th>
<th>Wheat 2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extension</td>
<td>1.4791 (0.5732)</td>
<td>1.4731 (0.5703)</td>
<td>0.3600 (0.1742)</td>
<td>0.3718 (0.1754)</td>
</tr>
<tr>
<td>Literacy</td>
<td>-0.0011 (0.1138)</td>
<td>-0.0140 (0.1134)</td>
<td>0.2131 (0.0760)</td>
<td>0.1761 (0.0786)</td>
</tr>
<tr>
<td>Roads</td>
<td>5.43e-07 (3.07e-06)</td>
<td>2.19e-07 (3.03e-06)</td>
<td>-7.12e-06 (1.48e-06)</td>
<td>-7.87e-06 (1.58e-06)</td>
</tr>
<tr>
<td>Bullocks/GCA</td>
<td>0.0166 (0.0113)</td>
<td>0.0162 (0.0113)</td>
<td>0.0018 (0.0064)</td>
<td>-0.0001 (0.0064)</td>
</tr>
<tr>
<td>Tractors/GCA</td>
<td>-0.0031 (0.0049)</td>
<td>-0.0028 (0.0049)</td>
<td>0.0226 (0.0060)</td>
<td>0.0230 (0.0060)</td>
</tr>
<tr>
<td>Wages</td>
<td>0.0001 (0.0033)</td>
<td>-0.0173 (0.0035)</td>
<td>-0.0187 (0.0060)</td>
<td>-0.0576 (0.0068)</td>
</tr>
<tr>
<td>Price/Pwheat(-1)</td>
<td>0.0029 (0.0029)</td>
<td>0.0024 (0.0023)</td>
<td>-0.0015 (0.0030)</td>
<td>-0.0027 (0.0032)</td>
</tr>
<tr>
<td>Price/Pwheat(-2)</td>
<td>0.0014 (0.0021)</td>
<td>0.0011 (0.0021)</td>
<td>0.0053 (0.0031)</td>
<td>0.0061 (0.0033)</td>
</tr>
<tr>
<td>Pmaize/Pgram</td>
<td>-0.0045 (0.0031)</td>
<td>-0.0034 (0.0031)</td>
<td>-0.0102 (0.0039)</td>
<td>-0.0097 (0.0040)</td>
</tr>
<tr>
<td>Psugar</td>
<td>-0.0001 (0.0023)</td>
<td>-0.0002 (0.0023)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>GIA/GCA</td>
<td>0.0191 (0.0071)</td>
<td>0.0187 (0.0071)</td>
<td>0.0178 (0.0057)</td>
<td>0.0166 (0.0056)</td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.0059 (0.0021)</td>
<td>0.0060 (0.0021)</td>
<td>0.0088 (0.0026)</td>
<td>0.0090 (0.0025)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0241 (0.0050)</td>
<td>-0.0220 (0.0050)</td>
<td>0.0124 (0.0035)</td>
<td>0.0171 (0.0036)</td>
</tr>
</tbody>
</table>

**Year dummies** mostly significant mostly significant mostly significant mostly significant

| Q     | 0.8347 (0.8374) | 0.8374 (0.8354) | 0.8354 (0.8564) |
| MI    | 3.9132 (3.9018) | 4.7963 (4.7871) |
| R²    | 0.0852 ---      | 0.1500 ---      |
| N     | 4711 4711       | 5500 5500       |

**NOTE:**
1. Q ~ $x_1^2$ under $H_0$: no serial correlation. Statistic is computed with level residuals.
2. For rice, MI ~ N(-0.24,0.0128) under $H_0$: no spatial correlation.
   For wheat, MI ~ N(-0.22,0.0127) under $H_0$: no spatial correlation.
   Statistic is computed with first-differenced residuals.
3. Standard errors are in parentheses.

122
<table>
<thead>
<tr>
<th></th>
<th>Rice OLS</th>
<th>Rice 2SLS</th>
<th>Wheat OLS</th>
<th>Wheat 2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extension</td>
<td>-0.6090</td>
<td>-0.7399</td>
<td>-1.0439</td>
<td>-1.0904</td>
</tr>
<tr>
<td></td>
<td>(1.1609)</td>
<td>(1.1607)</td>
<td>(1.2099)</td>
<td>(1.2120)</td>
</tr>
<tr>
<td>Roads</td>
<td>-7.07e-05</td>
<td>-7.76e-05</td>
<td>-6.76e-05</td>
<td>-6.54e-05</td>
</tr>
<tr>
<td></td>
<td>(2.86e-05)</td>
<td>(2.88e-05)</td>
<td>(3.35e-05)</td>
<td>(3.35e-05)</td>
</tr>
<tr>
<td>Literacy</td>
<td>-0.6947</td>
<td>-0.8795</td>
<td>0.8139</td>
<td>0.9303</td>
</tr>
<tr>
<td></td>
<td>(1.8980)</td>
<td>(1.9226)</td>
<td>(1.6264)</td>
<td>(1.6233)</td>
</tr>
<tr>
<td>Bullocks/GCA</td>
<td>-0.4955</td>
<td>-0.5024</td>
<td>-0.2072</td>
<td>-0.2007</td>
</tr>
<tr>
<td></td>
<td>(0.1085)</td>
<td>(0.1079)</td>
<td>(0.0893)</td>
<td>(0.0896)</td>
</tr>
<tr>
<td>Tractors/GCA</td>
<td>-0.1046</td>
<td>-0.0995</td>
<td>-0.0907</td>
<td>-0.0908</td>
</tr>
<tr>
<td></td>
<td>(0.0797)</td>
<td>(0.0791)</td>
<td>(0.0712)</td>
<td>(0.0711)</td>
</tr>
<tr>
<td>Wages</td>
<td>-0.0821</td>
<td>-0.4829</td>
<td>0.0299</td>
<td>0.1922</td>
</tr>
<tr>
<td></td>
<td>(0.0923)</td>
<td>(0.1028)</td>
<td>(0.0907)</td>
<td>(0.0937)</td>
</tr>
<tr>
<td>Price/Pwheat(-1)</td>
<td>0.3102</td>
<td>0.3058</td>
<td>0.1927</td>
<td>0.2013</td>
</tr>
<tr>
<td></td>
<td>(0.0602)</td>
<td>(0.0596)</td>
<td>(0.1122)</td>
<td>(0.1120)</td>
</tr>
<tr>
<td>Price/Pwheat(-2)</td>
<td>-0.0569</td>
<td>-0.0617</td>
<td>0.0237</td>
<td>0.0233</td>
</tr>
<tr>
<td></td>
<td>(0.0515)</td>
<td>(0.0516)</td>
<td>(0.0687)</td>
<td>(0.0684)</td>
</tr>
<tr>
<td>Pmaize/Pgram</td>
<td>-0.2612</td>
<td>-0.2366</td>
<td>0.0241</td>
<td>0.0251</td>
</tr>
<tr>
<td></td>
<td>(0.0914)</td>
<td>(0.0913)</td>
<td>(0.0583)</td>
<td>(0.0581)</td>
</tr>
<tr>
<td>Psugar</td>
<td>0.0305</td>
<td>0.0247</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.0566)</td>
<td>(0.0569)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIA/GCA</td>
<td>0.1057</td>
<td>0.0989</td>
<td>0.4094</td>
<td>0.4146</td>
</tr>
<tr>
<td></td>
<td>(0.0924)</td>
<td>(0.0918)</td>
<td>(0.0633)</td>
<td>(0.0636)</td>
</tr>
<tr>
<td>Rainfall</td>
<td>0.2923</td>
<td>0.2920</td>
<td>0.0501</td>
<td>0.0495</td>
</tr>
<tr>
<td></td>
<td>(0.0421)</td>
<td>(0.0416)</td>
<td>(0.0360)</td>
<td>(0.0361)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0020</td>
<td>0.0468</td>
<td>-0.0101</td>
<td>-0.0308</td>
</tr>
<tr>
<td></td>
<td>(0.0425)</td>
<td>(0.0429)</td>
<td>(0.0388)</td>
<td>(0.0391)</td>
</tr>
<tr>
<td>Year dummies</td>
<td>Some significant</td>
<td>Some significant</td>
<td>Some significant</td>
<td>Some significant</td>
</tr>
<tr>
<td>Q</td>
<td>0.8177</td>
<td>0.8442</td>
<td>0.7296</td>
<td>0.7045</td>
</tr>
<tr>
<td>MI</td>
<td>5.7026</td>
<td>5.6761</td>
<td>5.3289</td>
<td>5.3374</td>
</tr>
<tr>
<td>R²</td>
<td>0.2366</td>
<td>---</td>
<td>0.2808</td>
<td>---</td>
</tr>
<tr>
<td>N</td>
<td>4042</td>
<td>4042</td>
<td>4480</td>
<td>4480</td>
</tr>
</tbody>
</table>

**NOTE:**
1. \( Q \sim x_1^2 \) under \( H_0: \) no serial correlation. Statistic is computed with level residuals.
2. For rice, \( MI \sim N(-0.24,0.0128) \) under \( H_0: \) no spatial correlation.

For wheat, \( MI \sim N(-0.22,0.0127) \) under \( H_0: \) no spatial correlation.

Statistic is computed with first-differenced residuals.
3. Standard errors are in parentheses.
Table 4(a): Second-Stage Estimation: Learning Function [Rice]

<table>
<thead>
<tr>
<th></th>
<th>Restricted OLS</th>
<th>Unrestricted</th>
<th>Restricted 2SLS</th>
<th>Unrestricted</th>
<th>2SLS adjusted Restricted</th>
<th>2SLS adjusted Unrestricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_{12}$</td>
<td>0.2253</td>
<td>0.2320</td>
<td>0.2214</td>
<td>0.2240</td>
<td>0.2284</td>
<td>0.2283</td>
</tr>
<tr>
<td></td>
<td>(0.0114)</td>
<td>(0.0119)</td>
<td>(0.0116)</td>
<td>(0.0120)</td>
<td>(0.0119)</td>
<td>(0.0123)</td>
</tr>
<tr>
<td>$e_{22}$</td>
<td>0.2012</td>
<td>0.1993</td>
<td>0.2142</td>
<td>0.2139</td>
<td>0.2303</td>
<td>0.2284</td>
</tr>
<tr>
<td></td>
<td>(0.0109)</td>
<td>(0.0115)</td>
<td>(0.0110)</td>
<td>(0.0117)</td>
<td>(0.0112)</td>
<td>(0.0118)</td>
</tr>
<tr>
<td>$e_{32}$</td>
<td>0.1679</td>
<td>0.1616</td>
<td>0.1576</td>
<td>0.1522</td>
<td>0.1660</td>
<td>0.1595</td>
</tr>
<tr>
<td></td>
<td>(0.0104)</td>
<td>(0.0109)</td>
<td>(0.0103)</td>
<td>(0.0109)</td>
<td>(0.0106)</td>
<td>(0.0111)</td>
</tr>
<tr>
<td>$e_{42}$</td>
<td>0.1391</td>
<td>0.1366</td>
<td>0.1408</td>
<td>0.1405</td>
<td>0.1431</td>
<td>0.1409</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.0090)</td>
<td>(0.0090)</td>
<td>(0.0093)</td>
<td>(0.0089)</td>
<td>(0.0092)</td>
</tr>
<tr>
<td>$e_{52}$</td>
<td>0.0802</td>
<td>0.0770</td>
<td>0.0794</td>
<td>0.0755</td>
<td>0.0791</td>
<td>0.0734</td>
</tr>
<tr>
<td></td>
<td>(0.0101)</td>
<td>(0.0102)</td>
<td>(0.0100)</td>
<td>(0.0102)</td>
<td>(0.0097)</td>
<td>(0.0100)</td>
</tr>
<tr>
<td>$e_{62}$</td>
<td>0.0706</td>
<td>0.0682</td>
<td>0.0521</td>
<td>0.0514</td>
<td>0.0600</td>
<td>0.0565</td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td>(0.0123)</td>
<td>(0.0121)</td>
<td>(0.0123)</td>
<td>(0.0119)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>$e_{72}$</td>
<td>0.0471</td>
<td>0.0543</td>
<td>0.0547</td>
<td>0.0644</td>
<td>0.0621</td>
<td>0.0658</td>
</tr>
<tr>
<td></td>
<td>(0.0171)</td>
<td>(0.0180)</td>
<td>(0.0174)</td>
<td>(0.0182)</td>
<td>(0.0164)</td>
<td>(0.0171)</td>
</tr>
<tr>
<td>$e_{82}$</td>
<td>0.3330</td>
<td>0.3299</td>
<td>0.3405</td>
<td>0.3336</td>
<td>0.3153</td>
<td>0.2990</td>
</tr>
<tr>
<td></td>
<td>(0.0272)</td>
<td>(0.0285)</td>
<td>(0.0272)</td>
<td>(0.0293)</td>
<td>(0.0277)</td>
<td>(0.0297)</td>
</tr>
<tr>
<td>$e_{92}$</td>
<td>0.0112</td>
<td>0.0119</td>
<td>0.0255</td>
<td>0.0280</td>
<td>0.0372</td>
<td>0.0405</td>
</tr>
<tr>
<td></td>
<td>(0.0142)</td>
<td>(0.0145)</td>
<td>(0.0143)</td>
<td>(0.0148)</td>
<td>(0.0145)</td>
<td>(0.0149)</td>
</tr>
<tr>
<td>$e_{102}$</td>
<td>0.0112</td>
<td>0.0049</td>
<td>-0.0074</td>
<td>-0.0160</td>
<td>-0.0163</td>
<td>-0.0195</td>
</tr>
<tr>
<td></td>
<td>(0.0130)</td>
<td>(0.0133)</td>
<td>(0.0131)</td>
<td>(0.0140)</td>
<td>(0.0133)</td>
<td>(0.0140)</td>
</tr>
<tr>
<td>$e_{112}$</td>
<td>-0.0400</td>
<td>-0.0576</td>
<td>-0.0456</td>
<td>-0.0620</td>
<td>-0.0457</td>
<td>-0.0542</td>
</tr>
<tr>
<td></td>
<td>(0.0117)</td>
<td>(0.0123)</td>
<td>(0.0117)</td>
<td>(0.0123)</td>
<td>(0.0115)</td>
<td>(0.0120)</td>
</tr>
<tr>
<td>$e_{122}$</td>
<td>-0.0456</td>
<td>-0.0500</td>
<td>-0.0528</td>
<td>-0.0558</td>
<td>-0.0518</td>
<td>-0.0514</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0108)</td>
<td>(0.0106)</td>
<td>(0.0113)</td>
<td>(0.0099)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>$e_{132}$</td>
<td>-0.0581</td>
<td>-0.0540</td>
<td>-0.0551</td>
<td>-0.0499</td>
<td>-0.0549</td>
<td>-0.0489</td>
</tr>
<tr>
<td></td>
<td>(0.0108)</td>
<td>(0.0114)</td>
<td>(0.0108)</td>
<td>(0.0116)</td>
<td>(0.0101)</td>
<td>(0.0106)</td>
</tr>
<tr>
<td>$e_{142}$</td>
<td>-0.0377</td>
<td>-0.0354</td>
<td>-0.0277</td>
<td>-0.0247</td>
<td>-0.0313</td>
<td>-0.0261</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
<td>(0.0136)</td>
<td>(0.0126)</td>
<td>(0.0134)</td>
<td>(0.0120)</td>
<td>(0.0126)</td>
</tr>
<tr>
<td>$e_{152}$</td>
<td>-0.0189</td>
<td>-0.0140</td>
<td>-0.0228</td>
<td>-0.0147</td>
<td>-0.0209</td>
<td>-0.0118</td>
</tr>
<tr>
<td></td>
<td>(0.0177)</td>
<td>(0.0187)</td>
<td>(0.0181)</td>
<td>(0.0190)</td>
<td>(0.0162)</td>
<td>(0.0168)</td>
</tr>
<tr>
<td>$e_{162}$</td>
<td>0.0024</td>
<td>0.0028</td>
<td>0.0025</td>
<td>0.0026</td>
<td>0.0024</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0010)</td>
<td>(0.0009)</td>
<td>(0.0010)</td>
<td>(0.0009)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>$e_{172}$</td>
<td>most insign</td>
<td>most insign</td>
<td>most insign</td>
<td>most insign</td>
<td>most insign</td>
<td>most insign</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.5166</td>
<td>0.5060</td>
<td>0.5080</td>
<td>0.5014</td>
<td>0.4984</td>
<td>0.5061</td>
</tr>
<tr>
<td></td>
<td>(0.0266)</td>
<td>(0.0281)</td>
<td>(0.0267)</td>
<td>(0.0286)</td>
<td>(0.0275)</td>
<td>(0.0293)</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.3604</td>
<td>0.3455</td>
<td>0.3641</td>
<td>0.3477</td>
<td>0.3137</td>
<td>0.3026</td>
</tr>
<tr>
<td></td>
<td>(0.0175)</td>
<td>(0.0175)</td>
<td>(0.0172)</td>
<td>(0.0176)</td>
<td>(0.0175)</td>
<td>(0.0186)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0011</td>
<td>0.0006</td>
<td>0.0011</td>
<td>0.0006</td>
<td>0.0006</td>
<td>-0.0003</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
<td>(0.0005)</td>
<td>(0.0004)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>Q</td>
<td>-0.0209</td>
<td>-0.0171</td>
<td>-0.0211</td>
<td>-0.0172</td>
<td>-0.0238</td>
<td>-0.0222</td>
</tr>
<tr>
<td>MI</td>
<td>0.7290</td>
<td>0.7712</td>
<td>0.7149</td>
<td>0.7518</td>
<td>0.6426</td>
<td>0.6806</td>
</tr>
<tr>
<td>R^2</td>
<td>0.8855</td>
<td>0.8872</td>
<td>0.8960</td>
<td>0.8976</td>
<td>0.8831</td>
<td>0.8845</td>
</tr>
<tr>
<td>LR</td>
<td>83.4167</td>
<td>---</td>
<td>84.5415</td>
<td>---</td>
<td>70.8225</td>
<td>---</td>
</tr>
<tr>
<td>N</td>
<td>5632</td>
<td>5632</td>
<td>5632</td>
<td>5632</td>
<td>5632</td>
<td>5632</td>
</tr>
</tbody>
</table>

NOTE: $Q \sim \chi^2$, under $H_0$, no serial correlation. MI $\sim N(-0.53,0.013)$ under $H_0$, no spatial correlation. LR $\sim \chi^2_{30}$ for restricted model. Standard errors are in parentheses.
Table 4(b): Second-Stage Estimation: Learning Function [Wheat]

<table>
<thead>
<tr>
<th></th>
<th>Restricted</th>
<th>OLS Unrestricted</th>
<th>Restricted</th>
<th>2SLS Unrestricted</th>
<th>2SLS adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{1t}$</td>
<td>0.2279</td>
<td>0.2283</td>
<td>0.2371</td>
<td>0.2375</td>
<td>0.2432</td>
</tr>
<tr>
<td></td>
<td>(0.0078)</td>
<td>(0.0079)</td>
<td>(0.0078)</td>
<td>(0.0079)</td>
<td>(0.0081)</td>
</tr>
<tr>
<td>$\epsilon_{2t}$</td>
<td>0.2280</td>
<td>0.2277</td>
<td>0.2131</td>
<td>0.2118</td>
<td>0.2124</td>
</tr>
<tr>
<td></td>
<td>(0.0081)</td>
<td>(0.0082)</td>
<td>(0.0082)</td>
<td>(0.0083)</td>
<td>(0.0085)</td>
</tr>
<tr>
<td>$\epsilon_{3t}$</td>
<td>0.1786</td>
<td>0.1776</td>
<td>0.2011</td>
<td>0.2000</td>
<td>0.1969</td>
</tr>
<tr>
<td></td>
<td>(0.0078)</td>
<td>(0.0079)</td>
<td>(0.0079)</td>
<td>(0.0080)</td>
<td>(0.0082)</td>
</tr>
<tr>
<td>$\epsilon_{4t}$</td>
<td>0.1601</td>
<td>0.1618</td>
<td>0.1557</td>
<td>0.1553</td>
<td>0.1534</td>
</tr>
<tr>
<td></td>
<td>(0.0083)</td>
<td>(0.0083)</td>
<td>(0.0082)</td>
<td>(0.0083)</td>
<td>(0.0085)</td>
</tr>
<tr>
<td>$\epsilon_{5t}$</td>
<td>0.1070</td>
<td>0.1076</td>
<td>0.0951</td>
<td>0.0968</td>
<td>0.0872</td>
</tr>
<tr>
<td></td>
<td>(0.0095)</td>
<td>(0.0095)</td>
<td>(0.0091)</td>
<td>(0.0092)</td>
<td>(0.0092)</td>
</tr>
<tr>
<td>$\epsilon_{6t}$</td>
<td>0.1021</td>
<td>0.1062</td>
<td>0.0947</td>
<td>0.1035</td>
<td>0.1024</td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td>(0.0112)</td>
<td>(0.0105)</td>
<td>(0.0107)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>$\epsilon_{7t}$</td>
<td>0.0909</td>
<td>0.0929</td>
<td>0.0915</td>
<td>0.0950</td>
<td>0.0875</td>
</tr>
<tr>
<td></td>
<td>(0.0146)</td>
<td>(0.0148)</td>
<td>(0.0134)</td>
<td>(0.0139)</td>
<td>(0.0131)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{1t-1}$</td>
<td>0.1132</td>
<td>0.1096</td>
<td>0.1433</td>
<td>0.1238</td>
<td>0.1205</td>
</tr>
<tr>
<td></td>
<td>(0.0502)</td>
<td>(0.0518)</td>
<td>(0.0471)</td>
<td>(0.0504)</td>
<td>(0.0486)</td>
</tr>
<tr>
<td>$\epsilon_{2t-1}$</td>
<td>0.0670</td>
<td>0.0717</td>
<td>0.0523</td>
<td>0.0582</td>
<td>0.0687</td>
</tr>
<tr>
<td></td>
<td>(0.0139)</td>
<td>(0.0141)</td>
<td>(0.0135)</td>
<td>(0.0140)</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>$\epsilon_{3t-1}$</td>
<td>0.0992</td>
<td>0.0900</td>
<td>0.0173</td>
<td>0.0202</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>(0.0150)</td>
<td>(0.0154)</td>
<td>(0.0139)</td>
<td>(0.0144)</td>
<td>(0.0142)</td>
</tr>
<tr>
<td>$\epsilon_{4t-1}$</td>
<td>-0.0080</td>
<td>-0.0091</td>
<td>-0.0114</td>
<td>-0.0083</td>
<td>-0.0111</td>
</tr>
<tr>
<td></td>
<td>(0.0125)</td>
<td>(0.0127)</td>
<td>(0.0131)</td>
<td>(0.0136)</td>
<td>(0.0133)</td>
</tr>
<tr>
<td>$\epsilon_{5t-1}$</td>
<td>-0.0584</td>
<td>-0.0569</td>
<td>-0.0624</td>
<td>-0.0602</td>
<td>-0.0618</td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td>(0.0124)</td>
<td>(0.0116)</td>
<td>(0.0120)</td>
<td>(0.0119)</td>
</tr>
<tr>
<td>$\epsilon_{6t-1}$</td>
<td>-0.0649</td>
<td>-0.0637</td>
<td>-0.0730</td>
<td>-0.0686</td>
<td>-0.0735</td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td>(0.0113)</td>
<td>(0.0102)</td>
<td>(0.0106)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>$\epsilon_{7t-1}$</td>
<td>-0.0610</td>
<td>-0.0565</td>
<td>-0.0683</td>
<td>-0.0579</td>
<td>-0.0610</td>
</tr>
<tr>
<td></td>
<td>(0.0124)</td>
<td>(0.0127)</td>
<td>(0.0114)</td>
<td>(0.0120)</td>
<td>(0.0115)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon_{8t-1}$</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0007)</td>
<td>(0.0007)</td>
<td>(0.0007)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>$\epsilon_{9t-1}$</td>
<td></td>
<td></td>
<td>most insig</td>
<td>most insig</td>
<td>most insig</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>0.5939</td>
<td>0.5936</td>
<td>0.5677</td>
<td>0.5835</td>
<td>0.5876</td>
</tr>
<tr>
<td></td>
<td>(0.0512)</td>
<td>(0.0527)</td>
<td>(0.0482)</td>
<td>(0.0515)</td>
<td>(0.0499)</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.1353</td>
<td>0.1297</td>
<td>0.1390</td>
<td>0.1274</td>
<td>0.1163</td>
</tr>
<tr>
<td></td>
<td>(0.0324)</td>
<td>(0.0331)</td>
<td>(0.0285)</td>
<td>(0.0314)</td>
<td>(0.0305)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Q</td>
<td>-0.0065</td>
<td>-0.0057</td>
<td>-0.0075</td>
<td>-0.0069</td>
<td>-0.0075</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0003)</td>
<td>(0.0004)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>MI</td>
<td>0.2289</td>
<td>0.2113</td>
<td>0.2501</td>
<td>0.2326</td>
<td>0.2955</td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.0241)</td>
<td>0.9441</td>
<td>0.9444</td>
<td>0.9339</td>
</tr>
<tr>
<td>R²</td>
<td>38.0635</td>
<td>11.8725</td>
<td>58.7541</td>
<td>32.7874</td>
<td>61.5712</td>
</tr>
<tr>
<td>N</td>
<td>5491</td>
<td>5491</td>
<td>5491</td>
<td>5491</td>
<td>5491</td>
</tr>
</tbody>
</table>

**Note:** Q $\sim x^2_1$ under $H_0$: no serial correlation. MI $\sim N(-0.54,0.013)$ under $H_0$: no spatial correlation. LR $\sim x^2_{30}$ for restricted model. LR $\sim x^2_{7}$ for partially restricted model. Standard errors are in parentheses.