Chapter 1

INTRODUCTION

1.1 Background

1.1.1 Importance of foodgrain production

Agriculture is considered the foundation of the national economy in China. It accounted for 51.7 per cent of the net output value of agriculture and industry and 44.9 per cent of national income in 1983 (MAAF 1983). It also provided light industry with large amounts of raw materials which made up about 70 per cent of the total output value of light industry in the same year (Lu 1984). Further, about 40 per cent of China's export goods are from the agricultural sector (MAAF 1983).

Experience in the last 37 years indicates that economic growth in China hinges largely on the satisfactory performance of the agricultural sector. For example, from 1979 to 1982, China's agricultural production showed an annual average growth rate of 7.5 per cent, which enabled industry to obtain an annual growth of 7.2 per cent, GNP 7.5 per cent, and national income 6.3 per cent (Lu 1984). On the other hand, the well-known 'three-year-disaster period' in agriculture (i.e., from 1959 to 1961) resulted in a sharp drop in GNP from 267.9 billion* *yuan*† in 1960 to 197.8 billion *yuan* in 1961, and a further drop to 167.7 billion *yuan* in 1962 (MAAF 1983).

*1 billion = 100 million.

†As at the end of June, 1989, 3.7 *yuan* approximately equalled US$1.

The significance of agriculture directly implies the critical position of foodgrain production in China since agricultural production basically means foodgrain production. As a matter of fact, some people (even economists) in China have on occasion confused foodgrain production with agriculture, which led to a debate among economists on the concepts of *big agriculture* and *small agriculture*[‡] in the early 1980s. The dominant position of foodgrain production in agriculture can be characterised by the following aspects: (a) Foodgrain production valued at 109.11 billion *yuan* in 1983, 61 per cent of the total agricultural production value (based on 1980 prices); (b) Area sown for grain (grain will be used as the synonyrn of foodgrain in this thesis) in 1983 was 17.1 billion *mu* 79 per cent of the total area sown; (c) At least 80 per cent of the 300 million agricultural labour force was engaged in grain production (MAAF 1983, Anon. of People's Daily 1986b).

To emphasise the importance of foodgrain production in China more directly, it is worth noting the following points: (i) China has to feed about 23 per cent of the world's population using only 7 per cent of the world's arable land. The population of China was 1024.95 million in 1983 with a conservative prediction of 1115.94 million in 1990 (Li 1983), while the arable land is only about 21.2 billion *mu* with a prediction of a 82.5 million *mu* decrease by the year 1990 (Bureau of Land Administration of MAAF 1984); (ii) Grain in China accounted for more than 85 per cent of both total energy intake and protein intake (Zhang 1984). Thus grain is nearly synonymous with food consumption; (iii) About 80 per cent of the population depend on agriculture which is dominated by foodgrain production (Bodin 1985); (iv) As a political matter, grain production directly relates to social stability; (v) More importantly, China is not doing much better than self-sufficiency in grain. The future of China's foodgrain production is not optimistic at all (see section 1.1.3).

From a global perspective, China produced 20.3 per cent of the world's grain and was the largest producer of rice and wheat, and the second largest producer of maize, sorghum and millet in 1984 (Bodin 1985). Its shares in rice, wheat, maize, sorghum and millet were,

[‡]Big agriculture is defined as the totality of farming, forestry, animal husbandry and side-production (which includes rural industry) and fishery, while small agriculture simply indicates foodgrain production or farming (which includes cash crop production in addition). In this thesis, agriculture always represents so-called *big agriculture* .

respectively, 38.1, 16.8, 15.3 and 15.4 per cent of the world total (Stone and Zhong 1985). Although China has been basically self-reliant in food and would like to continue in this way, given the absolute figures on population and foodgrain production, its importance in the world grain market cannot be overemphasised.

Furthermore, China is one of the largest traders in the world grain market. Its wheat imports of 276 billion *jin* in 1980/81 were second only to Russia's, accounting for 14.8 per cent of total world imports (Tang 1982). Chinese rice exports averaged around 10 per cent of the world total in the late 1970s and were 10.31 per cent in 1980, exceeded only by those of the United States (24.0 per cent) and Thailand (21.6 per cent) (Kueh 1984). Overall, China is a net importer of grain, ranking among the five largest traders in the world grain market. As stated by Tang (1982), China looms large in the global food situation; the country's role in world food trade has been increasing, and fairly small relative changes in its food balance either as short-run manifestations or as long-term trend tendencies translate into large absolute magnitudes with serious impact on the global picture.

Looking to the future, demand calls for producing some 268 billion *jin* of feedgrain, 580 billion *jin* of foodgrain and 176 billion *jin* of grain for other uses in the year 2000, but area sown for grain will decrease about 150 million *mu* by that time (CAAS 1984, Bureau of Land Administration of MAAF 1984). Gains in yields due to improved varieties, capital construction and fertilizer use are expected to be limited (Bodin 1985). Also the initial incentive effect on production stemming from the introduction of the agricultural production responsibility system (APRS) has leveled off. This can be seen by the fact that a lot of farmers were not willing to invest in foodgrain production (Commentator of People's Daily 1986b). The irrational price structure coupled with power decentralization will lead to a further drop of area sown for grain. Worse still, population control may become more difficult under the new economic system. These considerations do not make for an encouraging picture of China's foodgrain production for the next 15 years.

Increasing grain trade is being predicted not only because of the gap between domestic demand and supply, but also due to the substantial price disparity between wheat and rice in the international market. Therefore, one has little doubt to say that China will retain its position as one of the world's largest traders.

1.1.2 Development of foodgrain production in China

Prior to the foundation of the People's Republic of China in 1949, China was a semifeudal and semicolonial country with a backward agriculture. Total grain production was 2264 billion *jin* in 1949. The average yield was 137 *jin* per *mu*. More than 70 per cent of the agricultural land was owned by less than 10 per cent of the rural population. Foodgrain production was mainly by tenants. According to a survey, tenants only had an average of 15.6 *mu* of land split into several unconnected blocks (Zhao,Zhu and Liu 1980). Farmers were unable to have any capital inputs, for no surplus farming products were left after the high rent, normally more than 60 per cent of the physical product (Zhao, Zhu and Liu 1980).

To confiscate land from the rich for redistribution among the poor, land reform first started in 1928 in the revolutionary bases—areas under Communist control. Though it was a political campaign and stopped several times from 1928 to 1949, it more or less certainly promoted foodgrain production(Zhao, Zhu and Liu 1980). Nation-wide land reform was undertaken between 1949 and 1952, which entitled 300 million farmers to own 700 million *mu* land. The significance of the land reform is reflected in the sharp increase of yield from 137 *jin* per *mu* in 1949 to 176 *jin* per *mu* in 1952. During this period the total value of agricultural production rose by 48.5 per cent, or 14.1 per cent annually (MAAF 1983).

1953 to 1957 was the first five-year-plan period, during which cooperative transformation of agriculture was implemented by three steps: (a) initiation of mutual aid groups; (b) establishment of elementary agricultural production cooperatives; and (c) foundation of advanced agricultural production cooperatives. Until the last step, the agriculture system was not fully socialist in nature.

Step (a) only involved labour–labour exchange or labour–tool exchange among farmers. The characteristic of step (b) was the co-existence of collective management and private ownership. The use of land, tools and animal power were arranged by the cooperative and the owners obtained dividends according to their shares of land and other property. Step (b) brought agriculture to semi-socialism status. When step (c) was completed, the individual family lost the ownership of land and the collective took over the private animals and major tools by a kind of redemption.

The cooperative transformation of agriculture had a positive effect on foodgrain production. Total grain output increased from 3377 billion *jin* in 1953 to 3900 billion *jin* in 1957, and the average yield rose by 18.7 *jin* per *mu* over this period (MAAF 1983).

The infamous 'great leap forward' commenced in late 1958, immediately after the cooperative transformation. In the same year, 23 630 communes emerged. The 'big rice bowl', 'equalitarianism', 'blind commanding' and statistical falsification were the products of the mistaken campaign. This was a political pursuit rather than an economic motivation. The campaign, though stopped in the winter of 1960, severely affected China's foodgrain production. Total foodgrain output dropped to 2870 billion *jin* and yield to 156 *jin* per *mu* in 1960 (MAAF 1983). The 'three-year-disaster period', partly attributable to the 'great leap forward', worsened foodgrain production further. Not until 1964 did grain yield recover to the level of 209 *jin* per *mu*.

The following 'cultural revolution' (1966–76) brought the whole nation to the edge of collapse both politically and economically. Surprisingly, agricultural production, especially foodgrain production, still grew at a moderate pace. Total grain production increased by 1446 billion *jin*, equivalent to an annual growth rate of 2.95 per cent. There were at least three reasons for the unusual growth: (1) Farmers in general remained in rural areas and did not stop farming even though they also fought each other sometimes; (2) Governments devoted most of their efforts to develop foodgrain production. During this period, many huge agriculturally-oriented projects were completed (e.g., the Hui river project, the North Plain irrigation system) and large amounts of non-agricultural land were converted to foodgrain production. The once world-known 'Dazhai' was an example to convert naked mountains to paddy fields; (3) New technology was more quickly adopted under the radical commanding system. The adoption of high-yielding varieties (including semi-dwarf wheat, rice and hybrid rice, maize) was probably faster in China than in any other countries.

The agricultural production responsibility system (APRS) was introduced in 1978. This initiated the national economic reform. The introduction of APRS can be called the 'Second Land Reform' in the sense that this reform entitled individual families to receive a certain amount of land from the collective, and collective farming was actually dissolved. The land is still owned by the collective and so farmers have to deliver agricultural products to the government according to a contract. The APRS was initially introduced in poorer areas and later became the uniform farming system. The foodgrain production sector has performed exceptionally well since the initiation of APRS. This is shown by the increases in total grain output and yield (see Table 1.1).

Table 1.1: China's Foodgrain Production: 1976–85

Year	Area Sown (10^4 *mu*)	Yield (*jin/mu*)	Output (10^8 *jin*)
1976	181115.00	316.15	5726.00
1977	180600.00	313.12	5655.00
1978	180881.00	347.36	6283.10
1979	178894.10	369.70	6613.00
1980	174708.50	364.30	6364.40
1981	172436.50	376.97	6500.40
1982	170093.50	415.60	7068.50
1983	171070.80	452.80	7745.50
1984	169325.90	481.10	8146.10
1985	163267.70	464.40	7582.16

Source: Rural Statistical Yearbook of China (various issues).

Besides the institutional changes and favorable weather, the price adjustment in 1979 with grain price up by 21 per cent contributed greatly to this achievement. Now the family is the basic production unit and is the dominant supplier of grain. On the other hand, the compulsory procurement system introduced in 1955 was replaced by the contract purchasing system in 1985. Although the contract purchasing system is more consistent with the goal of using economic means to manage the economy, it certainly threatens the grain supply to the state (Editorial of Economic Daily 1986).

1.1.3 Future growth: a question of grain 'self-sufficiency'

Several studies have attempted to make projections of China's grain production. Among these are Kueh (1984), Noh (1983), Kilpatrick (1982) and Tang and Stone (1980), who made relatively extensive quantitative analysis. According to Kueh, the most plausible conclusion would be that, by 1990, China could still be self-sufficient in grain, with a possible 19.2 billion *jin* net export or a net import of 179.6 billion *jin*. By the year 2000, depending on the meat–grain substitution realized, China might become a net grain exporter with the export amount ranging from 15.8 billion *jin* to 485.6 billion *jin* . However, a more liberal income and consumption policy could conversely put China in a position as a net importer, the amount of imports possibly reaching as much as 612.8 billion *jin*. A somewhat less optimistic prediction is that energy and financial limitations would prevent sufficient increase in the inputs needed to produce enough grain domestically and large amounts of grain, possibly 1264 billion *jin*, would have to be imported for use as livestock feed (Kilpatrick 1982). Worse still, Noh (1983) found that China might become a rice importer in the middle of the 1990s. He predicted that almost 20 billion *jin* of rice would be imported in the year 2000. The total amount of grain imported could be 458 billion *jin* in 1990 and 714 billion *jin* in the year 2000. The median scenario of Tang and Stone's projection called for the total demand for grain to reach 10760 billion *jin* and total supply 10480 billion *jin* (excluding soybeans) at the turn of the century.

In the author's opinion, the projections by Mock and Kilpatrick are unrealistic and will not be commented on in detail. The projection by Kueh is over-optimistic primarily due to the facts that: (1) There is not enough evidence to support the assumption of a 2.67 per cent annual growth rate during the period 1980–2000. Reductions in area sown and the multi-cropping index may well be more substantial than Kueh's unknown expectation; (2) The favorable weather cycle anticipated by Kueh (1986) may be offset by an unfavorable weather trend (Zhu 1984); (3) The estimated income elasticities of food consumption given by Kueh, especially that for meat which is very influential, are too low. A log-linear function, which fits the income and meat consumption data (1952–82) well, shows that the income elasticity of demand for meat is 1.38 with a T-ratio of 3.69. This estimate of elasticity is 32 per cent higher than that used by Kueh; (4) The projection of foodgrain production of 10880 billion *jin* in the year 2000 is far more than the production target set by the Chinese government, which is 9600 billion *jin*, and it is expected to be difficult to achieve (Commentator of People's Daily 1986b).

As far as Tang and Stone's projections are concerned, the projected figures may look reasonable. However, there are two drawbacks in his approach. First, they rested their arguments on the 1975–85 Ten–Year Plan, which has been proven too ambitious and unrealistic. Second, there are substantial discrepancies between the official data and the data used by Tang and Stone.

In recent years, Chinese economists have tried to forecast the total supply and demand of grain in China. One notable study was initiated by the Rural Policy Research Office under the Central Committee Secretary's Department in 1982. At least 12 ministries or commissions and 5 universities participated in this study. It is expected that the final results will not be published. However, available information indicates that an expected shortage of 240 -- 300 billion *jin* grain is estimated for the year 2000.

It is clear that, unless a fundamental breakthrough in agricultural technology happens, China will be no more than self-sufficient in grain as it enters the 21st century. This is well-recognized by the Chinese government (Commentator of People's Daily 1986b, c, Anon. of Guangming Daily 1986, Anon. of Economic Daily 86d, Yao 1986).

1.1.4 Instability of China's foodgrain production

If, in the future, China could maintain self-sufficiency in grain with acceptable levels of imports, it would be fluctuations in foodgrain production rather than the production itself, which would be of main concern to the Chinese and other national governments.

Total production, area sown and yield (1949–86) are plotted on Figure 1.1, which clearly demonstrates significant year-to-year changes. Using 1952–57 and 1979–83 data, Stone and Zhong (1985) found that the coefficients of variation of total foodgrain production, area sown and average yield increased by 286, 343 and 183 per cent respectively over the two periods.

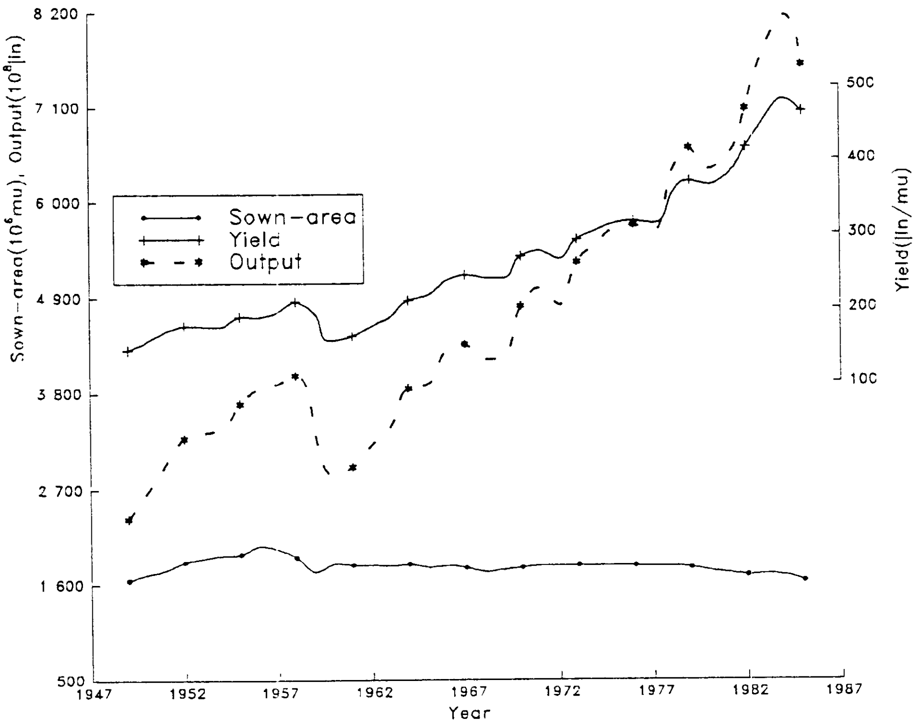

Figure 1.1: Chinese Foodgrain Output, Area Sown and Yield: 1949–86

According to MAAF statistics, year-to-year variations in national foodgrain production range from a 14.1 per cent or 404 billion *jin* increase between 1951–52 to a 15.6 per cent or 530 billion *jin* drop between 1960–61. The most striking effects of this instability were a net population loss, estimated at minimum to be 13.48 million during the 1960–61 famines (SSB 1985), and extensive grain loss in 1982–83 due to the sudden jump in production and improper storage and transportation facilities.

As for any agricultural and developing country, instability in foodgrain production may lead to suffering amongst most of the population. Hunger is the first and most common phenomena when foodgrain production varies downwards. In China, farmers suffer more than anybody else. They cannot benefit much from a good harvest because of low grain prices and because they traditionally sacrifice food consumption in order to prevent starvation in bad years. Recalling that foodgrain production is still the dominant source of agricultural income, the rigid price policy followed in China increases fluctuations in farmers' incomes, which directly influence farmers' lives. Furthermore, when foodgrain production decreases, even those farmers with money cannot obtain sufficient food since priority of grain supply from the state reserve is normally given to city and town residents.

Treating instability as an indicator of risk and assuming that farmers in China are risk-averse, it is to be expected that fluctuations in foodgrain production will reduce the average return to resources, and further reduce the supply of grain (Quiggin and Fisher 1989, Fraser 1989). This is particularly true when the agricultural production system becomes more market-oriented. On the other hand, the introduction of the APRS in 1978 and the abolition of the grain procurement system in 1985 have had much effect on the increment of foodgrain production instability since the early 1980s. This is evidenced by an unusually large number of public speeches and publications by senior leaders and the government media on grain matters (State Council 1986, Editorial of Economic Daily 1986, 1987, Commentator of People's Daily 1986a, b, c, Commentator of Economic Daily 1986, 1987, Yao 1986, Anon. of Economic Daily 1986a, Anon. of People's Daily 1986a, Anon. of Economic Daily 1986c, d, Anon. of Guangmin Daily 1986). The vice-premier, Yilin Yao (1986), wrote an article titled 'Sparing no effort on solving grain problems'. On the 13th of December 1986, the

Economic Daily cited vice-premier Jiyun Tian's speech as the first item on the cover page, which was sub-titled: 'Cherish land resources and stabilize area sown for foodgrain' (Anon. of Economic Daily 1986b).

On the world market, as China participates more in international trade and as import–export power is further decentralized, the global grain situation will be more strongly affected by changes in China's foodgrain production. For example, following the large grain surplus in 1982–83, US farm exports to China dropped to 48 billion *jin* in 1984–85 from an amount of 198 billion *jin* in 1981–82 (Bodin 1985). It seems reasonable to predict that in the long run, China could be one of the largest traders in the world grain market because its potential comparative advantage is very attractive (Yang 1986) and because advantages due to any price disparity between rice and wheat are likely to be exploited more extensively in the future.

Consequently, research on China's foodgrain production, and more specifically on its instability, is of great significance, not only to China but also in a global context.

There are not many quantitative studies of China's foodgrain production and even fewer, if any, focusing on the causal factors of foodgrain production instability. It is commonly accepted that weather disturbances dominate the sources of yield variations. The author also believes that political intervention must have played a determining role in changing the area sown for grain. Market mechanisms have probably only had limited influence on China's foodgrain production in recent years. The international market has had very little effect, if any, on domestic grain supply. Nevertheless, it must be said that the last two factors will possibly affect the instability of foodgrain production to an increasing extent in the near future.

The problems of foodgrain production and its instability have received considerable attention by the Chinese government since 1949, and even earlier. Remarkable work has been done on promoting high yield and stable production. This includes large-scale water-control projects, irrigation system construction, better-variety breeding and labour-intensive farming techniques. Although it is still hard to assess the efficiency of these efforts economically, to a certain extent they do help in stabilizing national grain production.

1.2 Literature Review

1.2.1 Definition of instability

There appears to be no uniform definition of instability in the literature and it is difficult to propose one which can be distinguished from such concepts as risk and uncertainty. Surprisingly, some authors have not discussed the meaning of instability in their papers. Some might have thought it was unnecessary because their analytical framework would offer a hint on the definition(e.g. Lawson 1974, Casley, Sinaika and Sinha 1974, Girão, Tomek and Mount 1974). Others implicitly used instability as a synonym of risk, e.g., Lin (1977), Hazell (1982). However, since both instability study and risk analysis utilize statistics, whatever methodology is being employed cannot imply the meaning of concepts under consideration unless they are explicitly defined. On the other hand, risk and instability are indeed two different concepts despite the fact that they are closely related. The difference between risk and instability was addressed by Quiggin and Anderson (1979) in the context of decision-making. They argued that decisions made without knowing what realization a random variable would take are subject to both risk and instability, while decisions made with this knowledge are subject to instability only.

Instability is defined here as the extent of variation of a time series around its predictive trend. The characteristics of instability according to this definition include: (1) Instability is objective and measurable. To measure instability, time series data are essential; (2) No matter what measure one is going to use, it must be computed around the trend of the time series. The trend removed is assumed to be predictable.

There are at least two distinctions between risk and instability. First, risk always relates to the future, while instability is only relevant to the past. Second, risk analysis always involves personal or subjective beliefs, while instability study need not do so. Loosely speaking, perceived unknown instability should be viewed as risk and risk in the past should be renamed as instability. It is common practice to use information from instability analysis to make inferences about risk.

To differentiate uncertainty from instability, the definition given by Roumasset (1979) is followed. He describes uncertainty as a state of mind in which the individual perceives alternative outcomes to a particular action. In this sense of being a state of mind, uncertainty is not quantitatively measurable.

In their review paper, Offutt and Blandford (1983) made an effort to distinguish variability from instability. They argued that instability is only a part of variability, the unacceptable variability. Of course, to each person the unacceptable level is different. Consequently, instability is subjective and a unique definition does not exist. The author believes that there is no clear-cut difference between variability and instability and so the distinction is unnecessary. Throughout this thesis, variability will be used as a synonym of instability.

It is worth noting an important contribution by Gelb (1979). He introduced spectral analysis into instability study. This technique made it possible to separate and distinguish fluctuations in a time series for different durations, and to measure them by one model. He provided an index $F(d_0, d_1)$ on a time series X, which measured the variation of the frequency components of X with duration $[d_0, d_1]$. Gelb implied that this index was a precise definition of instability in the sense that, unlike other indices, it enabled cross-study comparability and indicated clearly the type of fluctuation being measured. The present author appreciates the flexibility of $F[d_0, d_1]$ as a measure of instability, but he does not accept the 'precise definition' as a definition. It may be better to view this index as a distinctive measure of instability.

1.2.2 Measurement of instability

Research on agricultural production instability is relatively scarce except in the context of price stabilization or food security policy. Most instability measures have been developed for export studies. They are generally expressed as indices. At least 16 distinct such indices can be found in the development literature alone. It is not necessary to review them in this thesis. The more commonly used measures are variance, standard error, coefficient of variation and its square.

There are two ways to classify instability measures . First, a measure may be dimensionless (relative measure) or unit-dependent (absolute measure). For instance, the coefficient of variation (CV) and its variants belong to the former group, while variance and mean absolute deviation belong to the latter group. Second, if a measure is constructed to calculate the magnitude of individual deviations from a predicted trend, it is an arithmetic measure; alternatively, a measure may be obtained based on the ratios of the actual observations to the predicted values. This measure is referred to as a geometric measure. A geometric measure can be made equivalent to an arithmetic one through normalization (Brodsky 1980). There is no link between these two kinds of classifications.

The choice of measure is very important since different measures often give substantially different results . Policy suggestions drawn by employing one measure may be significantly altered should another indication of instability be used (Lawson 1974, Gelb 1979, Offutt and Blandford 1983, Newbery and Stiglitz 1981). In choosing a measure, the objective of the study plays a determining role. For example, relative measures must be used if a cross-sectional comparison of instability is under consideration. Whether detrending is needed or what form it should take will also depend on the purpose of the study. For a detailed assessment of various instability measures, readers are referred to Offutt and Blandford (1983) and Lawson (1974).

1.2.3 Methods of instability decomposition

Some symbols have to be defined before proceeding further. Variance is represented by $Var(\cdot)$, covariance by $Cov(\cdot)$, standard deviation by $\sigma_{\cdot}$, correlation coefficient by $\rho_{\cdot\cdot}$ and expectation operator by $E(\cdot)$, where $\cdot$ indicates relevant variable. A bar is also used to indicate mean operation. Thus, $\bar{X} = E(X)$.

To the author's best knowledge, variance decomposition is the only technique yet used to decompose instability into quantitative components. It has been most widely applied to identities and in a few cases to simple regression equations. There are two kinds of identities: additive or multiplicative. The additive identity takes the following form

$$Y = \sum_{i=1}^{n} X_i \tag{1.1}$$

and the variance decomposition of equation (1.1) is shown below:

$$Var(Y) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j). \tag{1.2}$$

However,

$$Cov(X_i, X_j) = \rho_{X_i,X_j}\sigma_{X_i}\sigma_{X_j}. \tag{1.3}$$

Therefore, equation (1.2) can be written as

$$Var(Y) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \rho_{X_i,X_j}\sigma_{X_i}\sigma_{X_j}. \tag{1.4}$$

In words, the variability of Y can be partitioned into the variability of X_i and the interaction of pairs of variables ($Cov(X_i, X_j)$) or dependence between pairs of variables (ρ_{X_i,X_j}).

A multiplicative identity may be considered as

$$Y = \prod_{i=1}^{n} X_i \tag{1.5}$$

and an exact variance decomposition formula can be found in Goodman (1960, 1962) when n takes any integer. A useful version of equation (1.5) is when $i = 2$, i.e.,

$$Y = X_1 X_2. \tag{1.6}$$

According to Goodman (1962), the variance of equation (1.6) can be expressed as

$$\begin{aligned} Var(Y) &= E^2(X_1)Var(X_2) + E^2(X_2)Var(X_1) \\ &+ 2E(X_1)E(X_2)Cov(X_1, X_2) - Cov^2(X_1, X_2) \\ &+ 2E(X_1)E_{12} + 2E(X_2)E_{21} + E_{22}, \end{aligned} \tag{1.7}$$

where

$$E_{ij} = E\left[(X_1 - \bar{X}_1)^i (X_2 - \bar{X}_2)^j\right].$$

Economists often use assumptions to simplify equation (1.7). This probably results in either inaccuracy or bias, which may lead to faulty policy implications. The degree of inaccuracy or bias depends on how rational the assumptions are. For some agricultural economists (e.g., Burt and Finley 1968), the central limit theorem lends them the support to propose normal distributions of X_i. In this case, equation (1.7) reduces to

$$\begin{aligned} Var(Y) &= E^2(X_1)Var(X_2) + E^2(X_2)Var(X_1) \\ &+ 2E(X_1)E(X_2)Cov(X_1, X_2) - Cov^2(X_1, X_2) - E_{22}. \end{aligned}$$

If X_1 can be assumed to be expectation and variance independent of X_2, or vice versa — a weaker assumption than stochastic independence, then equation (1.7) can be written as

$$Var(Y) = E^2(X_1)Var(X_2) + E^2(X_2)Var(X_1) + Var^2(X_1)Var^2(X_2). \tag{1.8}$$

A more complicated identity may take either of the following forms:

$$Y = \sum_i \left(\prod_{j=f(i)} X_j \right) \tag{1.9}$$

$$Y = \prod_i \left(\sum_{j=f(i)} X_j \right) \tag{1.10}$$

where $j = f(i)$ means that j is a function of i.

Variances of equations (1.9) and (1.10) can be readily derived given the previous discussion, though they are rather complicated. Equation (1.10) rarely appears in economic quantities[§]. Hazell (1982, 1985) applied variance decomposition to national foodgrain production which was the sum of products of area sown and the yield of each individual region. He may have been the first economist to have decomposed the variance of an equation, like equation (1.9), without any assumptions as to the distributions of any of the relevant variables.

So far we have only discussed identities which can hold by definition. An important alternative to such identities is due to Powell (1960) and Piggott (1978). To decompose the variability of total gross revenue into demand and supply components, Piggott used simple linear regression, with price as the only explanatory variable, to estimate demand and supply

[§]But, if X_is could be estimated by ARIMA (integrated autoregressive-moving-average) models, this identity would be fairly common.

functions and solved for the equilibrium price and quantity. These two equilibrium values were then expressed as functions of demand and supply shifts. An alternative identity was constructed by equating total gross revenue with the product of expressions of equilibrium price and quantity. By applying the variance decomposition technique to this new identity, instability of gross revenue could be attributed to contributions made by demand, supply and their inseparable interactions. In comparison with the approach using the identity that revenue equals price times quantity, This approach can be used to identify the *root* sources of revenue instability, since price and quantity changes are often determined by movements of supply and demand in a competitive market. This contribution by Piggott and Powell made it possible to apply variance decomposition to identities not intuitively set by definition.

However, identities seldom exist among the variables under study, e.g. production and weather factors, demand and income. Where an identity cannot be defined or cannot be constructed as by Piggott (1978), it is necessary to use econometric techniques to estimate the equation. The principle of econometric modelling used here is that we should specify the model as simply as possible, which implies linearity in parameters and a minimum number of variables. This is justified on the grounds that variance decomposition of complex equations is too involved, if possible, and more importantly that it may make the interpretation of the results confusing, if not impossible. Of course, simplicity is a principle of any econometric modelling although this does not mean reducing the quality of the model.

Following is a linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \hat{\mu}, \tag{1.11}$$

where $\hat{\mu}$ is the estimated well-behaved residual and β_i are the estimated coefficients assumed for the moment to be non-stochastic parameters. The variance of Y in equation (1.11) is then

$$Var(Y) = \sum_{i=1}^{n} \beta_i^2 Var(X_i) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \beta_i \beta_j Cov(X_i, X_j) + Var(\hat{\mu}). \tag{1.12}$$

If all the independent variables are orthogonal, equation (1.12) becomes

$$Var(Y) = \sum_{i=1}^{n} \beta_i^2 Var(X_i) + Var(\hat{\mu}).$$

An application of variance decomposition with a multiple regression equation was carried out by Firch (1977). There are two unsolved problems associated with using econometric models in variance decomposition. The first, which was raised by Offutt and Blandford (1983), concerns multicollinearity among the independent variables and correlations between the residual and the independent variables. The other has to do with interpretations of β_i^2 as weight in the variance equation (Piggott 1978).

A further, and perhaps more important consideration may be added to the above. The parameters of an empirical regression equation are, in fact, stochastic and a more comprehensive accounting of uncertainty must also involve the variances and covariances of parameters (β_i). Erroneous conclusions could be drawn if the β_is were treated as constant rather than random variables.

Techniques other than variance decomposition are also available in instability study, though they are less developed and of limited use. Nevertheless, it is relevant to review some of them.

Brook, Grilli and Waelbroeck (1977) developed a framework for empirical analysis of the revenue effects of a buffer-stock price stabilization scheme on developing countries. Their conclusion depended on the dominant source of price instability in supply or demand. If supply variation is the main source, price stabilization will benefit exporting nations. On the other hand, if demand dominates as a source of price variability, importing countries will gain from price stabilization.

To identify the main contributor to price instability, a simple regression equation was postulated as

$$Q = \beta P + e,$$

where

Q: observed quantity expressed as deviations from its trend [so $E(Q) = 0$];

P: observed price expressed as deviation from its trend [so $E(P) = 0$];

e: residual comprising effects of variables other than P.

However,

$$\begin{aligned}\beta &= (P'P)^{-1}P'Q \\ &= \frac{\sum PQ}{\sum P^2} \\ &= \frac{Cov\,(P,Q)}{Var\,(P)}\end{aligned}$$

so

$$Cov(P,Q) = \beta\, Var(P).$$

Since $Var(P) \geq 0$, a positive (negative) β or $Cov(P,Q)$ implies that demand (supply) is the main source of price variability.

To enable analysis of multiplier effects in the context of instability study, a dynamic but stable system of equations must be used. Denoting an impact multiplier by π_{ij} (i: equation index; j: variable index), and the first order moving range of X_{tij} by M_{tij}, a coefficient (μ_{ij}) can be constructed for X_{ij}, i.e.,

$$\mu_{ij} = \pi_{ij} \sum_{t=2}^{T} |M_{tij}|\,.$$

By simply comparing μ_{ij} for every i, i.e., ranking $\mu_{i.}$, it can be shown which X_j, for any Y_i, is important in terms of explanation of Y_i's movement (Offutt and Blandford 1983). However, μ_{ij} does not indicate whether the exogenous variable has a stabilizing or a destabilizing effect on Y_i and μ_{ij} is subject to changes in the measurement units of X_{ij}.

1.3 Introduction to the Research

1.3.1 Objectives and methods

The objectives of the present study are to:

(1) Reveal the distribution and main sources of instability in China's grain production and measure the magnitude of the instability contributed by related factors.

(2) Measure changes in the instability of Chinese foodgrain production over time.

(3) Assess institutional effect on China's foodgrain production variability.

(4) Quantify the relationship between changes in China's foodgrain production variability and changes in inputs.

(5) Provide policy-makers with relevant suggestions.

In attempting objectives (1) to (3), single variable measures of instability will be employed extensively to make cross-region and cross-crop comparison of variabilities. Also, variance decomposition based on a defined identity will be used to separate the contribution to production instability made by variations in area sown and in yield. Decomposing the square of CV rather than the variance will be also attempted. By breaking relevant time series into two or more parts, changes in instability and in its components will be analysed.

Objective (4) can be fulfilled by estimating a stochastic production function with heteroscedastic components (Just and Pope 1978, Griffiths and Anderson 1982). An extension to the model considered by Griffiths and Anderson (1982) will be undertaken.

1.3.2 Distribution of grain production and scope of the study

As indicated by Figure 1.2, China is broadly divided into nine agricultural regions. Excluding regions 8 and 9, which are mainly engaged in animal husbandry, the so-called eastern part produces nearly 98 per cent of China's grain. Generally speaking, southern China grows most of the rice, while northern China is dominated by wheat. As an illustration, the south-east part (regions 5-7) grows more than 90 per cent of the country's rice, and about 85 per cent of the country's wheat is planted in the north-east part (regions 1-4). Using 1985 data, the proportion of total foodgrain output for each region¶ is presented in Table 1.2.

It is intended to include all the regions in this study. However, data are unavailable for Hebei, Jiangxi, Fujian, Menggu, Xizang (Tibet), Liaolin, Beijing and Yunnan. Also,

¶Prior to 1988, there were 22 provinces, 3 metropolitan cities and 4 autonomous regions in China with Taiwan excluded. For convenience, they are all referred to as regions in this thesis.

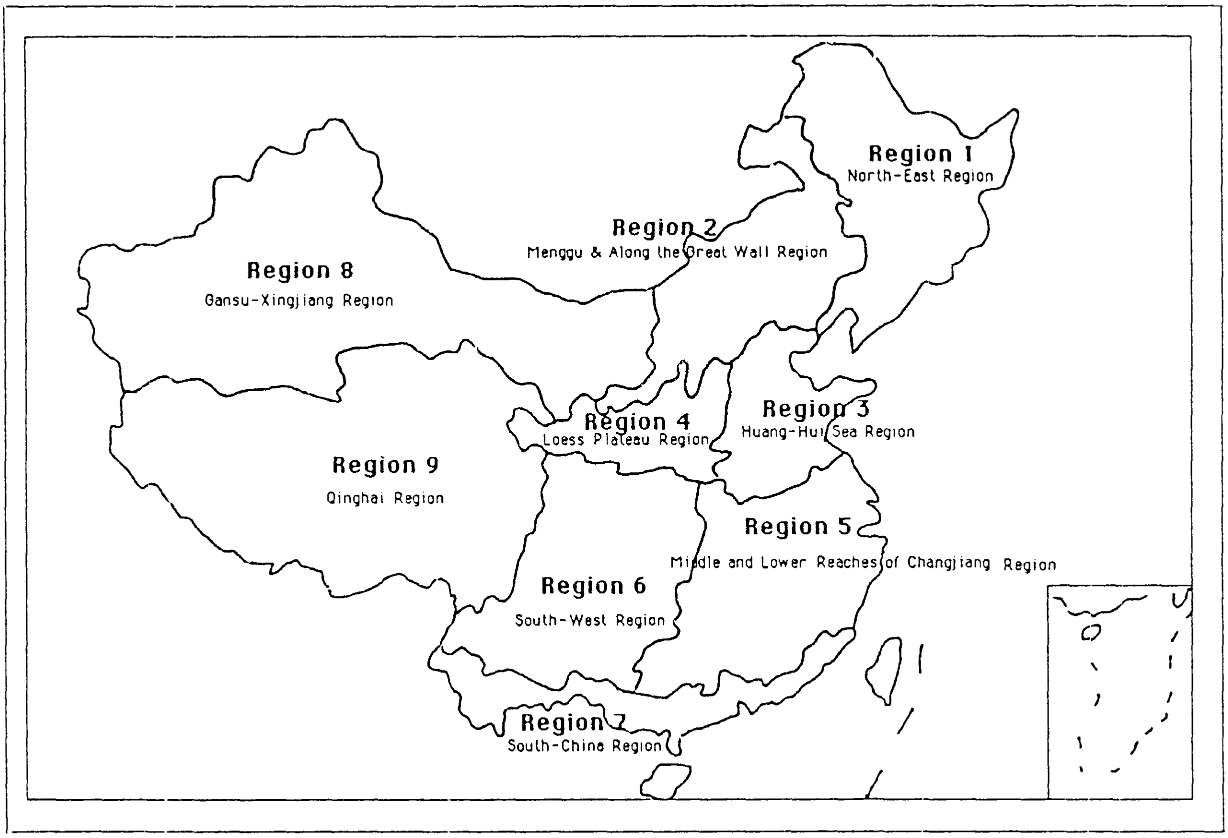

Figure 1.2: Agricultural Regions of China

Table 1.2: Composition of China's Foodgrain Production by Region: 1985

Region	Output (10^8 *jin*)	Share (%)	Area Sown (10^4 *mu*)	Share (%)
Beijing	43.94	0.58	767.10	0.47
Tianjin	28.10	0.37	668.90	0.41
Hebei	393.32	5.19	9739.10	5.97
Shanxi	164.54	2.17	4582.60	2.81
Menggu	120.82	1.59	5132.70	3.14
Liaoning	195.20	2.57	4334.30	2.65
Jilin	245.06	3.23	4925.20	3.02
Heilongjiang	286.00	3.77	10824.60	6.63
Shanghai	42.76	0.56	657.40	0.40
Jiangsu	625.30	8.25	9648.70	5.91
Zhejiang	324.26	4.28	4906.80	3.01
Anhui	433.60	5.72	8847.90	5.42
Fujian	158.88	2.10	2832.70	1.74
Jiangxi	306.70	4.05	5476.40	3.35
Shangdong	627.54	8.28	11976.40	7.34
Henan	542.10	7.15	13544.00	8.30
Hubei	443.22	5.85	7662.40	4.69
Hunan	502.86	6.63	7742.10	4.74
Guangdong	347.58	4.58	6698.20	4.10
Guangxi	223.42	2.95	5170.90	3.17
Sichuan	766.14	10.10	14083.50	8.63
Guizhou	119.00	1.57	3318.10	2.03
Yunnan	187.00	2.47	4977.80	3.05
Xizhang	10.62	0.14	291.10	0.18
Shaanxi	190.38	2.51	5948.40	3.64
Gansu	106.10	1.40	4162.30	2.55
Qinghai	20.06	0.26	579.90	0.36
Ningxia	27.90	0.37	975.70	0.60
Xingjinag	99.76	1.32	2792.50	1.71
China	7582.16	100.00	163267.69	100.00

Source: 1986 Rural Statistical Yearbook of China (1987).

Table 1.3: Composition of China's Foodgrain Production by Crop: 1985

Crop	Output (10^8 *jin*)	Share (%)	Area Sown (10^4 *mu*)	Share (%)
Rice	3371.38	44.46	48105.00	29.46
Wheat	1716.10	22.63	43827.00	26.84
Tubers	520.72	6.87	12858.00	7.88
Maize	1276.52	16.84	26541.00	16.26
Sorghum	112.18	1.48	2905.00	1.78
Millet	119.54	1.58	4978.00	3.05
Soybeans	210.00	2.77	11577.00	7.09
Other-grains	255.72	3.37	12477.00	7.64
China	7582.16	100.00	163268.00	100.00

Source: 1986 Rural Statistical Yearbook of China (1987).

data for some crops of those regions with available data can not be obtained. Therefore, depending on the crops, 12 to 21 regions will be included in this study. These regions cover about 80 per cent of total foodgrain production in China (refer to Chapter 2).

In discussing coverage of crops in this study, the Chinese definition of grain is followed, which includes potato (both white and sweet) and soybeans. Potato is still the main food in some parts of China and is converted to grain by the ratio of 5 : 1. Therefore, foodgrains consist of rice, wheat, maize, potato, soybeans, sorghum, millet and other-grains. It is noted that other-grains represent the difference between total foodgrain and sum of the seven crops just mentioned. Barley is included in other-grains in China due to its minimal output. The production percentages of the crops are shown in Table 1.3.

In summary, the study will cover rice, wheat, maize, potato, soybeans, sorghum, millet and other-grains, if regional data are required. In this case, 21 regions will be included. However, when national data are used, all crops and all regions will be covered in the analysis. The time period spans the 37 years from 1949 to 1985.

1.3.3 Data and Outline of the Study

There are two important problems associated with Chinese data, namely, reliability and source. Comparisons made by the author reveal that different sources gave remarkably different figures, especially for regional production data. For instance, the grain output for Hebei in 1957 given by MAAF is 1640 billion *jin*, but the figure is 202.6 billion *jin* according to CAAS.

The data problem has been a focal point for overseas economists and observers of the Chinese economy. They have been frustrated by the paucity or absence of official data release, particularly in the 1960s and 1970s. Thus, various kinds of estimations, even guesses, have been used to obtain Chinese statistics (Chen 1972, Cheng 1973, Wong 1973, Walker 1976). Among other problems is data reliability which has attracted much attention, especially following a limited outflow of Chinese information since 1978 (Tang and Stone 1980, Arid 1982, Emerson 1982, Hardt 1982, Stone and Zhong 1985). In brief, many believed that data from China were of a relatively low quality, and agricultural data worse than that for other sectors. However, after extensive checking and calculations, they also found that even the agricultural data were internally consistent and usable. A detailed treatment and more references on Chinese data can be found in Chen (1975) and Arid (1982).

Official reports have indicated the existence of falsification of statistics in China and the author is aware of the possibility that two sets of data might be kept by the government. Internal data (vs public data) certainly offer reliability to the best possible extent. It is extremely difficult to access the internal data, and even the public one if complete time series for longer periods are required.

The most annoying difficulty in studying the Chinese economy is the lack of quantitative data. While the preceding section provides an introduction to the data problem, a more detailed discussion on this topic will be presented in the next chapter.

The main task of chapter 3 is to rank variabilities of area sown, yield and production by crops and/or regions, which is followed by variance decomposition of total foodgrain production and CV^2 decomposition in chapter 4. These two chapters serve to reveal the sources and distribution of instability in foodgrain production in China.

In chapter 5, changes in the instability of China's foodgrain production will be examined with special references to institutional effect.

The relationship between changes in variability of production and changes in inputs can be best described by formulating so-called Just-Pope (1978) models. The estimations of the functions and their implications are presented in chapter 6.

This study concludes with the presentation of an epilogue in chapter 7.

Chapter 2

DATA

2.1 A Note on Chinese Data

Data problems are a major obstacle to study of the Chinese economy. There was no regular data release from China until the early 1980s and most of the official data initially available were either in percentage form or increment form. For example, area sown for wheat in China was 20 per cent of the total area sown for grains in 1959; the total grain output of Beijing in 1965 had risen by 40 billion *jin* from the previous year. Also, the Chinese data lately released do not constitute a complete series. In most cases, data are given for years 1952, 1957, 1965, the 1970s and the 1980s only. Finally, the releases seldom contain disaggregated regional data. For a large country like China, aggregate nation-wide data are of limited value for economic research, especially in agronomic studies.

Worse still, reliability of the data available is in question. In fact, some have replaced the word 'reliability' by 'usability' in discussing Chinese data issues. To ensure usability, internal consistency of the data on Chinese foodgrain production used in this study has been checked. The checking consisted of two steps: (a) to calculate the differences between the output and sown area of total foodgrain and the sum of outputs and sown areas of the individual crops for each region; the differences should not exceed what can be possibly caused by rounding errors; (b) to calculate the difference between production and the product of area sown and yield, and the difference between yield and the fraction of production over area sown for each crop of a region. If both differences are greater than unity, the corresponding figures are considered inconsistent. It was found that, except for some printing and data entry errors, the Chinese data on foodgrain production are in general consistent and usable.

2.2 Data Required for the Study

Three sets of data are required for this study, namely production data, weather data and data on foodgrain production inputs. The production data are composed of outputs and sown areas for different crops at the regional level. The weather data should comprise monthly rainfall and temperature records at station level (there are some 300 meteorological stations in China). Ideally, the inputs data should be at regional level and disaggregated by crops. Such disaggregated data, however, are not attainable for the whole period 1949–85. The inputs include chemical fertiliser, organic fertiliser, irrigation, labour, mechanisation and pesticide .

2.3 Problems with Chinese Data

2.3.1 Sources and availability of the data

It is noted that sources of Chinese data from 1979 onwards are no longer of serious concern to foreign scholars because the Chinese government unified the data release since then. Also, data released since 1979 are believed to be reasonably reliable and complete. Therefore, discussion in this section is confined to data for the period from 1949 to 1979.

As far as production data are concerned, they can be obtained from MAAF, SSB, CAAS or collected from individual regions. The author holds three sets of production data, which are, respectively, from CAAS, MAAF and individual regions. The coverage of the data from different sources in terms of the number of regions is presented in Table 2.1.

It can be seen that the MAAF data contain only total area sown and production for 29 regions and no disaggregated series by crops are available. The CAAS data only provide

Table 2.1: Coverage of Regions of the Production Data by Sources

	CAAS Data	MAAF Data	Regional Data
	.... number of regions		
Rice	29	na[1]	21
Wheat	26	na	21
Soybeans	25	na	21
Tubers	27	na	20
Maize	nu[2]	na	19
Sorghum	nu	na	16
Millet	nu	na	12
Other-grains	na	na	12
Foodgrain	26	29	20

[1] Data at regional level not available.
[2] Data at regional level too incomplete to be usable.
Sources: CAAS, MAAF and various regional reports.

relatively complete information on four crops. Although the data collected from each region (herein referred to as regional production data) do not cover as many regions as the CAAS data do, they cover more crops. Most of the missing regions in the regional production data produce insignificant amounts of foodgrain except Hebei, Menggu and Jilin. The regional production data offer complete crop coverage for 12 regions. If millet and other-grains could be disregarded, 16 regions would have complete crop coverage. If sorghum is further excluded from the list as an individual crop, 19 regions would have complete crop coverage.

Since the regional production data will be used for this study (see section 2.4.1 below), more detailed description of the data is given below.

The coverage of the regional data in terms of the national totals are reported in Tables 2.2 and 2.3 for selected years. The regions not covered by the regional production data but producing a significant amount, say, more than 5 per cent of the national crop in 1985 , include Hebei (8.7 per cent of wheat, 5.5 per cent of tubers, 10.3 per cent of sorghum, 10.6 per cent of maize and 24.4 per cent of millet), Menggu (7.9 per cent of sorghum and 13.1 per cent of millet), Jilin (8.6 per cent of soybeans, 12.4 per cent of maize, 10.1 per cent of

Table 2.2: Production Coverage of the Regional Data in National Totals by Crop

		1955	1960	1965	1970	1975	1980	1985
Rice								
China	(Mt)	78.1	59.6	87.7	110.0	125.6	139.9	168.5
21 Regions	(Mt)	64.6	47.1	74.0	88.2	106.3	114.5	139.2
%		82.8	79.0	84.3	80.2	84.6	81.9	82.6
Wheat								
China	(Mt)	23.0	22.1	25.2	29.3	45.3	55.2	85.8
21 Regions	(Mt)	20.2	18.8	23.0	25.6	39.2	47.6	75.0
%		88.1	85.1	91.3	87.5	86.6	86.3	87.4
Soybeans								
China	(Mt)	9.1	6.4	6.1	9.2	7.3	7.9	10.5
21 Regions	(Mt)	7.4	4.9	5.7	6.8	5.9	6.6	8.6
%		81.0	76.0	93.1	74.1	81.2	82.8	81.6
Tubers								
China	(Mt)	18.9	20.3	19.8	25.5	28.5	28.7	26.0
20 Regions	(Mt)	13.3	13.6	18.6	22.4	25.9	24.3	22.1
%		70.1	66.8	93.5	87.9	90.6	84.5	84.8
Maize								
China	(Mt)	20.3	na[1]	23.7	30.7	47.2	62.6	63.8
19 Regions	(Mt)	15.3	10.5	20.8	24.4	34.2	44.9	43.7
%		75.3	na	87.9	79.7	72.4	71.8	68.5
Sorghum								
China	(Mt)	10.3	na	7.1	8.2	10.8	6.8	5.6
16 Regions	(Mt)	7.8	2.5	5.4	6.1	7.3	4.9	4.0
%		75.6	na	75.8	74.3	68.2	72.6	70.7
Millet								
China	(Mt)	10.0	na	6.2	8.8	7.1	5.4	5.9
12 Regions	(Mt)	6.6	2.7	4.2	5.8	4.4	3.5	3.2
%		65.6	na	68.3	65.8	62.2	63.3	54.4
Foodgrain								
China	(Mt)	184.0	143.5	194.5	240.0	284.5	320.5	379.1
20 Regions	(Mt)	127.0	95.9	142.4	168.1	208.8	226.4	267.5
%		69.0	66.8	73.2	70.1	73.4	70.6	70.6

[1] Data not available.

Sources: Calculated from MAAF data and the regional data.

sorghum and 7.3 per cent of millet), and Jiangxi (8.7 per cent of rice). For convenience, the other eight regions, whose data are completely unavailable, will be referred to as the other-regions.

From Table 2.2, the regional production data covered some 80 per cent of the national outputs of rice, wheat, soybeans and tubers, about 70 per cent of the national outputs of maize and sorghum, and 60 per cent of the national output of millet. If MAAF data on Sichuan foodgrain production could be included for calculation (all data were available for Sichuan except the total foodgrain production), some 75 to 85 per cent of the national

Table 2.3: Sown-Area Coverage of the Regional Data in National Totals by Crop

		1955	1960	1965	1970	1975	1980	1985
Rice								
China	($10^6 ha$)	29.2	29.6	29.8	32.4	35.7	33.9	32.1
21 Regions	($10^6 ha$)	24.0	23.5	24.3	26.4	29.2	27.2	25.7
%		82.1	79.3	81.4	81.6	81.7	80.3	80.2
Wheat								
China	($10^6 ha$)	26.7	27.3	24.7	25.5	27.7	29.2	29.2
21 Regions	($10^6 ha$)	23.3	23.4	21.3	21.6	22.8	23.9	25.0
%		87.0	85.6	86.1	85.0	82.3	81.7	85.6
Soybeans								
China	($10^6 ha$)	11.4	9.3	8.6	8.4	7.0	7.2	7.7
21 Regions	($10^6 ha$)	9.2	7.4	7.0	6.4	5.7	6.0	6.4
%		80.3	79.0	81.9	76.2	81.3	82.7	83.4
Tubers								
China	($10^6 ha$)	10.1	13.5	11.2	10.3	11.0	10.2	8.6
20 Regions	($10^6 ha$)	8.1	10.5	9.4	9.0	9.1	8.5	7.2
%		80.6	77.5	84.0	88.0	82.9	83.3	83.9
Maize								
China	($10^6 ha$)	14.6	14.1	15.7	14.6	18.6	20.4	17.7
19 Regions	($10^6 ha$)	11.3	10.5	12.0	11.7	13.5	14.0	12.6
%		77.3	74.4	76.5	80.6	72.5	68.9	71.3
Sorghum								
China	($10^6 ha$)	8.1	3.9	6.1	4.8	4.7	2.7	1.9
16 Regions	($10^6 ha$)	6.1	2.7	4.3	3.6	3.0	1.9	1.3
%		75.4	67.9	69.4	75.4	63.3	69.4	68.8
Millet								
China	($10^6 ha$)	8.9	5.7	6.6	6.3	4.9	3.9	3.3
12 Regions	($10^6 ha$)	5.7	3.4	3.8	4.1	3.0	2.3	1.8
%		63.6	60.0	58.4	65.3	60.6	60.3	55.5
Foodgrain								
China	($10^6 ha$)	129.8	122.4	119.6	119.3	121.1	117.2	108.8
20 Regions	($10^6 ha$)	93.8	86.3	85.0	83.8	84.5	81.3	76.7
%		72.3	70.5	71.0	70.3	69.8	69.4	70.5

foodgrain output would be covered by the regional production data.

In terms of area sown, the regional production data covered 80 per cent of the national total for rice, wheat, soybeans and tubers, 70 per cent for maize and sorghum and 60 per cent for millet. If the MAAF data on Sichuan sown area for foodgrain could be used to replace the missing set, the 21 regions would share 80 per cent of the national area sown for foodgrain (see Table 2.3).

In brief, the regional production data obtained represent the dominant portion of Chinese foodgrain production.

Due to government restrictions and shortage of funds, the weather data could not be obtained from China and Hongkong as initially planned. Efforts were made to seek them through the US and Australian meteorological institutions. Fortunately, the records of World Monthly Surface Station Climatology (WMSSC) at the US National Center for Atmosphere Research (NCAR) were obtained through the Australian Bureau of Meteorology (ABM). This information was recorded in the order of longitudes and latitudes of stations all over the world, thus much effort had to be expended to extract the records of Chinese stations. After the extraction, it was found that the latest records are up to 1985 and that most of the stations had been moved several times for the past 30 years. The earliest records range from the 1840s to the 1960s. After careful selection, some 90 stations were chosen and further extractions were undertaken. The stations excluded are either located in non-agricultural areas or contain far insufficient data, especially from 1967 onwards. However, it was found that there were still too many missing values in the finally chosen data.

Attempting to obtain the regional input data by crop proved impossible due to time constraints and shortage of research resources. However, a set of survey data of inputs for rice, wheat and maize was made available to the author. The data cover only the period from 1980 to 1983 for 28 regions (with Xizang excluded). The data comprise output (*jin*), sown-area (*mu*), organic fertiliser cost (*yuan*), chemical fertiliser cost (*yuan*), machinery cost (*yuan*), irrigation cost (*yuan*), labour input (persondays) and other costs (*yuan*). Except for the area sown with modern cultivars, all the national input data were available to the author.

2.3.2 Missing values

As discussed in the preceding sections, all the three sets of the regional data required were available to varying degrees. There were still some missing values in the production, input and, particularly, weather data. Unfortunately, Chinese statistics never explain the nature of missing values. There are at least three reasons to explain the missing values in the production data set: (a) Absence of historical records. This mainly applies to the missing values between 1949 and 1952; (b) Malfunctioning of the statistical system which resulted in very unreliable data or absence of actual statistical work. When no evidence can be sought for adjusting the unreliable figures, they may be left over as missing values. This is the main cause for the missing values for the period between 1958 and the late 1970s; (c) Ignorance of minor crops. Since there were no uniform criteria for data collection or reporting, some regions did not separate some crops, like maize, sorghum and millet, from other-grains which are defined as the difference of total foodgrain less rice, wheat, maize soybeans, tubers, sorghum and millet, though other regions did. That is why there are no records of production of maize, millet and sorghum for some regions, e.g., Ningxia. Another example is the absence of tubers data for Xinjiang.

The number of missing values in the regional production data is reported in Table 2. 4. Missing values for other-grains are not counted because their existence mostly resulted from the absence of values for the crops as listed in Table 2.4. Overall, there were 159 observations missing and 18 sets of regional data were absent (see Table 2.4). Disregarding the absent data sets, the missing values occupied about 0.9 per cent of the total observations. The maximum number of missing values for a data set was 18, or 6 years' observations (area sown, yield and output figures for each year). This occurred for soybeans data of Hunan province. Since there are not many missing values, their estimation by interpolation, extrapolation or other techniques can be well-justified. The estimation will be undertaken in section 2.4.3.

There are far too many missing values in the weather data. There is no explanation for the missing values from NCAR and ABM. It is rather difficult and improper to estimate too many missing values. Thus, weather data will not be used in this study.

There are also a few missing values in the input data. Due to lacking formal techniques to estimate missing values in survey data and the necessity to employ balanced data in Chapter 6, 'best-guessed' numbers are used to replace those missing values. The guesses are based on (a) data from neighbour sample point; (b) data of previous or latter years if they are available; and (c) trend in the total regional input.

Table 2.4: Number of Missing Values in the Regional Production Data by Crops and Regions

	Rice	Wheat	Soy-beans	Tubers	Maize	Sor-ghum	Millet	Food-grain	Sum
Anhui	6	6	6	6	6	6	6	0	42
Hubei	0	0	0	0	0	0	0	0	0
Hunan	0	0	0	0	15	18	na[1]	0	33
Guangdong	0	0	0	0	na	na	na	0	0
Gansu	9	0	9	0	6	6	9	0	39
Guangxi	0	0	0	0	0	na	na	3	3
Guizhou	0	0	0	0	0	na	na	0	0
Heilongjiang	0	0	0	0	0	0	0	0	0
Henan	0	0	0	0	0	0	0	0	0
Jiangsu	0	0	0	0	0	0	0	0	0
Liaoning	0	0	0	3	0	0	0	0	3
Ningxia	3	3	3	3	na	na	na	3	15
Qinghai	0	0	0	0	0	0	na	0	0
Shaanxi	0	0	0	0	0	0	0	0	0
Sichuan	0	0	0	9	0	9	na	na	18
Shangdong	0	0	0	0	0	0	0	0	0
Shanghai	0	0	0	0	0	na	na	0	0
Shanxi	0	0	0	0	0	0	0	0	0
Tianjin	0	0	0	0	0	0	0	0	0
Xinjiang	0	0	0	na	0	0	6	0	6
Zhejiang	0	0	0	0	0	0	na	0	0
Sum	18	9	18	21	30	39	21	3	159

[1] Data not available.

Source: Calculated from the regional data.

2.3.3 Lack of unified statistical criteria

Another problem in compiling the production data was that each region presented its data in a different format. Some regions even did not give the total area sown and production of foodgrain, e.g., Sichuan, although some of these regions provided complete information on all crops and the summation of the individual crop information corresponds to the totals given by MAAF, e.g., Zhejiang. The individual crops listed in the reports ranged from 6 (Qinghai) to 17 (Hubei). Also some regions classified the production data into summer, autumn and winter crops but others did not. In reporting the total foodgrain production of a region, most regions included soybeans, while others like Hunan excluded soybeans.

The inclusion or exclusion of production of private plots and state farms' in the reported data raised another question. Several footnotes appearing in the regional statistical reports suggest that the state farms foodgrain production was included in the regional data. The fact that the state farm sector was administrated by the provincial or county government supports the same conclusion. As for production of private plots, scattered information in some of the regional data reports indicates that, prior to 1964, this production was included in the regional production. From 1965 onwards, some regions excluded this information, e.g., Shanghai; others excluded this production from the regional total and reported it separately as supplementary data. It is noted that both sown area and output of private plots represent 5 to 7 per cent of the regional total. Thus, it seems desirable to adjust the data by incorporating the information on production of private plots. However, due to the shortage of sufficient information, this adjustment is not possible.

When checking the measurement units used, it was found that about half of the regions used billion *jin*, ten thousand *mu* and *jin*/*mu* for output, area sown and yield, respectively. Other regions mostly used different units, some of them even using different units for different crops within one table.

2.3.4 Reliability of the data

The reliability of Chinese data was called into question initially because it was known that Chinese governments misled the public about the production progress by purposely publishing false statistics. This implies that governments perhaps kept the actual records. However, the author was aware of the fact that falsification in data reporting started from the grassroots — the production team. Thus, whether any actual records, in fact, existed became the fundamental point in discussing the reliability of Chinese data.

It is necessary to describe the causes of data falsification in order to assess the reliability of the data. To the author's understanding, two main factors motivated the data falsification, namely, avoiding agricultural tax and pursuing honourable reputation. Since agricultural tax in China was levied on total area sown, the farmers reported less land sown than the actual amount, and this occurred as early as the 1950s when the cooperative agriculture was established. Later, when production took the form of commune farming, the production team as the tax payer continued this practice. In some areas, the hidden land is called 'black land'. The existence of 'black land' was possible as there was no land survey information in China and the amount of land owned by each production team was estimated by the farmers or rural cadres. This is why the reported arable land of China (15 billion *mu*) is much less than the estimate of the arable land given by the Planning Commission of China (22 billion *mu*).

Data falsification for reputation was more extensively undertaken by commune and county governments. This pursuance reinforced the motivation of reporting less sown area in order to raise the yield figure. Greater claimed achievement in production might well attract more government funding and input allocations. In some areas, exaggeration became a necessary tool for protecting government officers' positions.

It is believed that governments at various level practised data exaggeration and that governments above county level often kept two or more sets of data, one of them free of exaggeration. The data not exaggerated were used for economic planning, and the other sets were reported to the higher levels of government and could be accessed by the public users.

If the arguments in the proceeding paragraphs hold, it would be true that there were no reliable data published prior to 1978. This conclusion needs to be modified for two reasons: (a) the large-scale sample survey on land resources conducted by the Planning Commission in the late 1970s and early 1980s would help to correct the arable land figure and then the production data; (b) the data re-adjustment organized by MAAF in 1979 would produce a set of data free of evident data falsification. Since this campaign started from the county government, the regional data obtained by the author and data from MAAF are seemingly of reasonable quality.

In summary, data were falsified at all government levels and two or more sets of data existed in China. One of the sets would be more reliable than others, though there are evidences to suspect the reliability of almost any set of Chinese agricultural data.

2.4 Solutions to the Data Problems

2.4.1 Choosing the right data source

As mentioned in the preceding sections and in Chapter 1, Chinese agricultural data are of low quality. However, it is believed, at least by the author, that data from MAAF and individual regions are more reliable than those from CAAS. This is because CAAS is not entitled to collect the production data, and the data obtained by CAAS can be from any source. It is known that falsification was practised in data reporting, and thus CAAS might have obtained false data should that be collected from local or regional governments. On the other hand, MAAF has the authority to compile the agricultural data, primarily for planning purposes, and it is less likely that local and provincial governments would provide false information to it. More importantly, when data were adjusted in 1979 for evident falsification, MAAF was the chief organizer. That is, MAAF data are free of any overt falsification. Due to the fact that adjusted data have not been made available for external use. CAAS data may not contain the necessary adjustment.

Although some differences do exist between MAAF and regional data, they are negligible and can be possibly attributed to rounding errors in the process of aggregation. The absence of disaggregated production data from MAAF prevents any detailed comparison between MAAF data and regional data. However, when 1949–78 data from the above-mentioned 21 regions were compared with the official release of 1979–86 data by MAAF, they appeared to be highly consistent. The statistical consistency between CAAS data and data from regions is tested in the next section.

2.4.2 Testing consistency between CAAS data and regional data

Although discrepancies between CAAS data and the regional data in some years are obvious, they may not be statistically significant. Assuming regional data are reasonably reliable, testing CAAS data against regional data is of importance for CAAS data users. More importantly, if the CAAS data are not significantly different from regional data, missing values in the regional data can be estimated from the available CAAS data.

Two kinds of statistical tests can be used to test consistency between different data sets, namely parametric tests and non-parametric tests. The latter have the advantage that no assumptions on the nature of the distribution of the relevant variables are needed. However, parametric tests are more powerful providing that the assumptions on the distributions of the relevant variables are valid. Since the Central Limit Theorem strongly supports the assumption of the normal distribution of foodgrain yields, sown areas and production, parametric tests are used.

Once the normal distribution is assumed, statistical tests should be used to test the equalities of the variances and then of the means. If both variances and means are not significantly different, it can be said that the two sets of data are from the same population. That is because two parameters, namely, mean and variance, completely specify a normal distribution. If variances are statistically equal and the means are not, simple regression could be used to adjust either of the data sets to equate their means should the reasons for the difference be found. When variances are significantly different, the t test for an equal mean is not applicable. In this case, same variance can be assumed if the relevant two samples are of equal size (Paradine and Rivett 1962, p. 118). However, the Fisher-Behrens test can be employed without assuming a common parent variance. Let m_i and s_i^2 denote mean and variance of sample i ($i = 1, 2$), this test requires calculation of θ and d, where, $\tan\theta = s_1/s_2$ and $d = (m_1 - m_2)/\sqrt{s_1^2 + s_2^2}$. The values of d at various significance level for $n_1, n_2 = 6, 8, 12, 24$, and ∞ are given in Fisher and Yates (1963) with intervals of 15^o in θ. Since samples are of equal size here, the Fisher-Behrens test will not be used to replace the t test with the assumption of equal variance.

Before proceeding to the tests, missing values in either CAAS or regional data should be handled. The principle adopted here is to exclude the observations in both sets of data if the values in any set for a particular year are missing.

Since CAAS data compiled by the author only include 5 categories (total foodgrain production, rice, wheat, soybeans and tubers) and regional data only cover 21 regions, the tests were conducted for some 315 paired sets of data (yield, sown area and output of 5 crops of 21 regions, that is, 315=3×5×21) often with different degrees of freedom for each pair. The significance levels for the tests are 5 per cent for means (t test) and 2 per cent for variances (F test) and in both cases the tests are two-tailed.

Using = to indicate equal and < or > to indicate that the mean or variance of CAAS data is less or greater than those of the regional data, the testing results are summarized in Table 2.5.

From Table 2.5, only 31 out of some 552 tests showed significant differences of means or variances between CAAS and regional data. This implies that CAAS data are rather consistent with the regional data. Since the regional production data were chosen for this study, no steps will be taken to adjust the CAAS data according to the test results unless the relevant CAAS data are used for estimating the missing values in the regional data.

2.4.3 Handling missing values and absent data set

The general consistency between CAAS data and the regional data offers one option to handle the missing values in the regional production data. The basic idea is to regress data from the regions on the corresponding CAAS set and the values predicted by the model can be used to fill in the missing values. However, in the case where tests show the equalities of

Table 2.5: Results of Testing for Equality of Variances and Equality of Means between CAAS Data and the Regional Data

	Area Sown		Yield		Output	
	Variances	Means	Variances	Means	Variances	Means
Anhui						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Hubei						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Hunan						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Guangdong						
Foodgrain	=	>	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	>	=	=	=	=	>
Soybeans	=	=	=	=	=	=

(Table continues on the next page)

(Table 2.5 continued)

	Area Sown		Yield		Output	
	Variances	Means	Variances	Means	Variances	Means
Gansu						
Foodgrain	>	=	=	=	=	=
Wheat	=	>	=	=	=	=
Rice	>	>	=	=	>	=
Tubers	=	=	=	=	=	=
Soybeans	>	=	=	=	=	>
Guangxi						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Guizhou						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	>	=	=	=	=	=
Heilongjiang						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Henan						
Foodgrain	<	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=

(Table continues on the next page)

(Table 2.5 continued)

	Area Sown		Yield		Output	
	Variances	Means	Variances	Means	Variances	Means
Jiangsu						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Liaoning						
Foodgrain	=	>	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Ningxia						
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Qinghai						
Wheat	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Shaanxi						
Foodgrain	=	=	=	=	=	=
Wheat	>	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Sichuan						
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	>	<	=	=	>
Soybeans	=	=	=	=	=	=

(Table continues on the next page)

(Table 2.5 continued)

	Area Sown		Yield		Output	
	Variances	Means	Variances	Means	Variances	Means
Shangdong						
Foodgrain	<	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Shanghai						
Foodgrain	>	=	>	=	>	=
Wheat	>	=	=	=	=	=
Rice	>	=	=	=	>	=
Tubers	=	=	<	=	=	<
Soybeans	<	=	<	=	<	<
Shanxi						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Xinjiang						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Soybeans	=	=	=	=	=	=
Zhejiang						
Foodgrain	=	=	=	=	=	=
Wheat	=	=	=	=	=	=
Rice	=	=	=	=	=	=
Tubers	>	=	=	=	=	=
Soybeans	=	=	=	=	=	=

Sources: Calculated from CAAS and the regional data.

both means and variances, the missing values in the regional production data can be simply replaced by the figures from the CAAS data if they are available. When values from both sources are missing, interpolation or extrapolation may be employed to infer the missing values based on the regional data.

Although there exist various methods of interpolation and extrapolation (Chinese Mineral University 1980), a polynomial equation of time (in year) is fitted into the available part of a regional production series and the predicted values by the equation can be used to replace the missing values. This method is preferred to other ones because of its simplicity and the availability of regression software. The degree of a polynomial is determined after inspecting the data plotted (Kelejian and Oates 1981). In doing so, subjective expectation of the range of the missing value(s) is incorporated. It must be pointed out that there is no justification in regard to the superiority of one method to another. Any attempt to estimate missing values is almost equally indefensible.

Referred to Table 2.4 and Table 2.5, the missing values in the regional production data were estimated differently as specified in Table 2.6.

The author is reluctant to estimate any absent data set. However, due to reasons mentioned elsewhere in this chapter, the absent total foodgrain data for Sichuan are replaced by those from MAAF. Also, the absent figures for Xinjiang tubers can be and were calculated as the difference between total foodgrain production and the sum of all the other crops. Once these values were estimated, missing values for other-grains were simply calculated.

2.5 The Finalized Data

The production data which will be used in this study are based on the regional production data, complemented by CAAS production data and the MAAF aggregated production data. The missing values in the regional data are either simply replaced by the corresponding CAAS or MAAF observations, or estimated by regression equations, interpolation and extrapolation depending on the result of consistency tests between the relevant data pairs and on the positions of the missing values in the data series. Unless otherwise noted, the

Table 2.6: Methods of Estimating Missing Values in the Regional Production Data

Region	Missing Values	Estimation
Anhui	All crops(1949–51)	Maize,sorghum and millet: extrapolation Other crops: replacement
Hunan	Maize(1966–70) Sorghum(1966–70, 1964)	Interpolation
Gansu	Rice, soybeans and millet (1967–69) Maize and sorghum (1967–68)	Rice and soybeans: prediction Other crops: interpolation
Guangxi	Maize(1949)	Extrapolation
Liaoning	Tubers (1949)	Replacement
Ningxia	All crops(1949) Foodgrain(1949)	Extrapolation
Sichuan	Tubers and sorghum (1949–51)	Tubers: prediction Sorghum:extrapolation
Xinjiang	Millet(1968–69)	Extrapolation

Sources: Calculated from CAAS and the regional production data.

Table 2.7: The Finalized Production Data

Crop	Number of Regions
Rice	21
Wheat	21
Maize	19
Tubers	21
Soybeans	21
Sorghum	16
Millet	12
Other-grains	12
Total Foodgrain	21

measurement units used in the data are *jin*, *jin/mu* and *mu*, respectively for output, yield and area sown. The final data comprise 1949 to 1985 time series of area sown, yield and output for 8 crops and more than 12 regions (see Table 2.7). These data are believed to be of the best possible quality, although further improvement can be made as information on private-plot production becomes available.

It must be noted that 1959-61 was an unusual period in China. The mistaken campaign of 'great leap forward', coupled with unfavourable climatic conditions, led to successive and dramatic decreases in foodgrain production over the years 1959-61, which has been referred to as 'Three-disaster-year'. Therefore, the period of 1959-61 will be excluded from this study unless otherwise specified.

Some survey data on inputs by crop are available only for the period from 1980 to 1983 for rice, wheat and maize. Regional input data not disaggregated by crop are available from 1980 to 1985. These data encompass chemical fertilizer, irrigation area, drainage capacity, electricity usage and mechanization, etc..