Multinational Corporations

The multinational corporation is a business organization whose activities are located in more than two countries and is the organizational form that defines foreign direct investment. This form consists of a country location where the firm is incorporated and of the establishment of branches or subsidiaries in foreign countries. Multinational companies can, obviously, vary in the extent of their multinational activities in terms of the number of countries in which they operate. A large multinational corporation can operate in 100 countries, with hundreds of thousands of employees located outside its home country.

The economic definition emphasizes the ability of owners and their managerial agents in one country to control the operations in foreign countries. There is a frequent confusion that equates the ability to control with the flow of capital across national borders. Since Hymer’s thesis (1976), it is recognized widely that capital flow is not the distinguishing characteristic of a multinational corporation (see International Business). Capital can flow from one country to another in expectation of higher rates of return. However, this flow may be invested in the form of bonds, or in equity amounts too insignificant to grant control to foreign owners. In this case, this type of investment is treated as a ‘portfolio’ investment. The central aspect of ‘direct investment’ is the ownership claim by a party located in one country on the operations of a foreign firm or subsidiary in another. The multinational corporation is, thus, the product of foreign direct investment that is defined as the effective control of operations in a country by foreign owners.

The economic definition, however, does not capture the importance of the multinational corporation as the organizational mechanism by which different social and economic systems confront each other. The multinational corporation, because usually it develops in the cultural and social context of one nation, exports its organizational baggage from one institutional setting to another. In this regard, it plays a powerful role as a mechanism by which to transfer organizational knowledge across borders. However, while being foreign implies that it might serve the valuable role of importing new practices, its foreign status also implies that its practices are likely to conflict with existing institutions and cultural norms. Moreover, since multinational corporations are often large, they pose unusual challenges to national and regional governments who seek to maintain political autonomy and yet are often anxious to seek the investment, technology, and managerial skills of foreign firms.

There are, thus, economic and sociological definitions of the multinational corporation that differ, and yet complement, each other. In the economic definition, the multinational corporation is the control of foreign activities through the auspices of the firm. In the sociological definition, the multinational corporation is the mechanism by which organizational practices are transferred and replicated from one country to another.

1. History

The multinational corporation is defined in some sense arbitrarily by where frontiers are drawn. In ancient Greece, these frontiers were the borders among city-states. In imperial Rome, the new administrative units of an expanding empire and its external boundaries defined the borders. Political borders correspond only approximately to the distribution of cultural and ethnic dispositions. The history of the multinational corporation is tied closely to the origins of trade in and between cultural communities, and these communities remain important in many sectors in the modern economy.

Early trade was often characterized by the difficulty of transacting across borders. Trading has always instigated from the unequal and varied distribution of resources across geographies. The trading of salt was an important factor in most continents and still can be found in rather ancient forms in east Africa. Towns, such as Salzburg, owed their origins to their fortuitous access to salt mines.

This unequal distribution incited traders to travel long distances and undergo unusual risks for the hope of gain. The dilemma is quite apparent. How can markets for trade develop between distant cultures? Indeed, there is good evidence that early trading was quite precarious. Brown (1969) notes that Hermes was not only the god of messengers but also of theft and
trade. He was often worshipped at borders, marking the place of trade. Herodotus describes the practice of the ‘silent trade’ of goods at boundary markings in which one party would deposit their goods at the border, the second party would later deposit a corresponding offer. The original party might accept this bid, take the goods, and leave their initial offer. Clearly, only a strong demand for unusual foreign goods could sustain such a treacherous institution of transaction.

Yet, trade persisted through the centuries. Indeed, the demand for foreign goods was so strong that incredible profits could be realized by international trade. Since transportation was poor, and technology to hold inventories rudimentary, an agent in the right spot at the right time could earn fortunes. And in this laid the great problem. For in times of great demand, prices could rise astronomically, as Braudel (1973) documents in his history of material life. With such distances between the agent and the principal in the exporting country, an agent could easily, and often did, disappear with the profits.

The early solutions were many and provide insight into one of the great properties of the multinational corporation: its ability to organize transactions within its own organizational boundaries. Great fairs were one solution, where principals met their customers in one place. But fairs were intermittent and often distant from final markets. The further development of Roman commercial law during the medieval and Renaissance periods strengthened the liability of the principal and also legal recourse in principal-agent conflict. However, the application of law between two parties in different legal jurisdictions has always been a difficult form of dispute resolution. In his study of the Jewish Magrebi traders of the twelfth century, Greif (1989) places considerable emphasis upon the value of membership in ethnic ‘trading communities.’ Since ethnic enclaves in foreign sites could monitor the activities of agents, principals could be informed of malfeasance. Agents could then be excluded from future contracts within the community, thus deterring dishonest behavior. Partnerships were also another way in which agents were also bonded to the firm.

However, another solution was the company, a word whose roots come from the Latin meaning ‘with bread’ which implies the familial origins of this business organization. The company permitted in time the investing of capital in its ownership by outsiders in promise of future dividends. The capital might be raised for a single voyage or, eventually, could be maintained in the company in the form of stock ownership. With the evolution of limited liability law in the nineteenth century, and the diminished role of the state in restricting the growth of the joint stock company, this organizational form expanded rapidly in many European countries as well as in the United States. Along with the growth of equity markets, the financial resources to invest overseas became more available due to advances in the banking system and also in bond markets. Indeed, the great capital needs of the railway industry created an international market for the sale and purchase of railroad system. However, international loans and bonds were risky and it was not uncommon for sovereign governments to default, including state governments within the United States.

The increasing wealth of western countries, along with constraints on the speed by which national industries could absorb new loans, encouraged massive foreign investments by the end of the nineteenth century and early twentieth century. This was a period of globalization in terms of the percentages of capital outflows to total capital accumulation. Great Britain is estimated to have exported some 25 percent of its capital prior to World War I and French capital exports were often greater. This capital went to countries hungry to finance their industrial expansion, including the newly industrializing countries of Russia and especially America.

The British were unusual in directing a considerable amount of their outward capital flows in the form of direct investments. A great deal of this investment went into their colonies. Prior to the mid-1800s, companies operating overseas engaged largely in wholesale operations; they did not run factories, operate mines, or own agriculture. The British direct investments were, thus, different than before, as British came to own and run local operations. Moreover, the British invested in South and North America.

The ‘free-standing’ company, as labeled by Wilkins (1988), was a peculiar form of investment that was quite prominent in many British colonies. This firm raised capital in the domestic financial market where its administrative office was located, but it operated no domestic activities. All operations were overseas. This form of company represents the advantage of western countries in being able to raise capital from efficient financial markets. However, it posed the same problem as facing the medieval agent relationship, namely, how could foreign investors believe that they would recover profitably their investments if the assets were located overseas, often in countries with very poor legal infrastructures? Wilkins found that oversight occurred through the executive boards of the companies, where prominent people located in the home market faced the loss of reputation if the company should prove to be dishonest.

The United States offers an unusual case. A rapidly growing country, it imported more capital than it
exported up to World War I. However, whereas it imported largely portfolio investments, its outward flows were dominated by foreign direct investments. In other words, American companies showed an early penchant for expanding overseas (Wilkins 1970). The Singer Company built within two decades of its founding a large factory employing thousands of workers in Scotland. Oil companies, Kodak, Westinghouse, Ford, and mining and agricultural companies all invested overseas. Companies in oil, mining, and agriculture often invested in poorer markets where there were resources to be found. These early investors were often involved inextricably in the politics of the foreign governments, and the American military itself intervened aggressively numerous times in the Caribbean, Mexico and Central America, and South America. As in the British case, the history of American direct investment occurred in the context of an expanding military presence of the home government. Moreover, since many of these countries were poor, the multinational corporations responded to the demands of the host nation, especially in the form of concessionary contracts, to provide public services, e.g., hospitals, roads, and power (Robinson 1964). This complicated legacy of the early history of the multinational corporation created hostility on the part of the local population that persisted throughout most of the twentieth century.

It is important to underscore that the multinational corporation usually evolved in the context of specific national institutions. As many others have pointed out, the multinational corporation is a growing firm whose organizational borders have spilled across borders. Moreover, since this large firm is usually tied to a larger domestic network of suppliers and customers, its expansion overseas is accompanied by the co-investments of these other members. This is a pattern seen in American investments in the United Kingdom in the 1950s and repeated by Japanese multinational corporations investing in the United Kingdom in the 1980s and 1990s (Dunning 1993).

Chandler (1990) noted that these multinational corporations reflected the national characteristics of management. In comparing the cases of the largest firms in the United Kingdom, Germany, and the United States, Chandler found that differences in managerial capabilities, reflecting national institutions, explain their success and failure patterns. He particularly criticized the managerial capabilities of British firms, a point not shared by some British historians (Hannah 1999). But more importantly, Chandler’s thesis assumed that size itself constituted the realization of scale and scope economies instead of the outcome of success and growth. This observation is especially important for understanding the lack of large multinational corporations in Italy or in Taiwan, both of which have very successful small firm economies but do not have multinational corporations comparable to other countries of similar levels of economic wealth. Yet, both countries are relatively wealthy and successful, and their many small companies have achieved high rates of exporting. Even in the case of the United States, the evidence implies that American firms, large and small, came to Europe, riding on the back of the national organizing principles of standardization in work methods (Kogut 1992). Chandler’s larger point of the effect of national systems on firm capabilities is largely accepted; his belief that large firms reflect better management because they achieve scale is disputed far more.

2. Organization of Multinational Corporations

The national origins of the multinational corporation influence their subsequent organizational evolution. In the United States, the great post-World War II expansion of multinational corporations coincided with the diffusion of their adoption of the multidivisional structure. Indeed, the refinement of ‘organizational technologies’ permitted American firms to manage their rapidly growing operations on a worldwide basis.

The young American multinational corporation began with little knowledge and experience of the foreign market. Initial foreign sales are exports executed within the existing organizational structure. These structures could be functional (e.g., organized by production and sales) or divisional (e.g., organized by geographic area or product division). As foreign sales increased, an international division was created to make the sale and to provide customer support. In consequence, the domestic organizational unit would sell to the international division its products at an internal ‘transfer price.’ As this price was set by the domestic market, it was unlikely to reflect the competitive conditions in the foreign country. Moreover, the products were often not designed to the tastes of foreign demand. Because of this internal conflict, the American multinational corporation would replace the international division by transferring responsibilities to area divisions (e.g., Europe, Asia, South America) or to product divisions (Stopford and Wells 1972). While these structures diminished the internal conflict, they ran the hazard that the divisional managers across areas did not cooperate or that managers across product divisions tended to fall back upon focusing on the home market.

European firms grew up in a different environment. Europe was a highly unstable continent during the twentieth century, with significant political and economic conflicts. Borders sometimes changed, and they also marked the necessity to pay commercial duties. The gradual creation of the European Union in the last third of the twentieth century eliminated tariffs and lessened national differences. This history marked their firms. The organizational technologies in most of
these countries were quite different to those found in the United States. Holding company structures were common, banks or other financial institutions played powerful roles in ownership and cross holdings, and firms were tied together in complex financial webs. These structures, often called ‘mother–daughter’ organizations, reflect the linguistic terms used in northern Europe to describe headquarters and subsidiary. These mother–daughter structures consisted often of a headquarters that held a portfolio consisting of ownership control over dozens of companies, many of them in the same sector but located in different countries. Reporting relationships among these companies and headquarters was less formal than found in American companies, and control was often exercised through the movement of managers who were lifetime employees. These structures were often seen as inefficient, especially compared to the size and organization of American multinationals that were investing rapidly in Europe at this time. European critics called for the restructuring of European firms and industry along American lines.

The structure of Japanese companies reflected also their historical origins. In Japan, the trading company performed an extraordinary role in the exports during the first two-thirds of the twentieth century. These companies started primarily as wholesalers, but over time developed their own industrial companies located in Japan and eventually overseas. Though the forced imposition of American antitrust law after World War II disrupted these patterns, trading companies and other affiliated company networks, called keiretsu, reconstituted the earlier pattern of constellations of domestic companies competing in home and, through their trading companies, in foreign markets. As in the American and European cases, Japanese firms varied in their internal structures, with functional and even factory-based organization being the predominant structures. Still, a dominant trend emerged over time as the role of trading companies declined, especially for the exportation of industrial and high-technology goods. This trend consisted of the appendage of an international division to the existing structure (Suzuki 1991).

These different national structures confronted an increasingly more globally integrated environment in the course of the twentieth century. This environment posed the classic organizational problem of balancing integration and differentiation. For the multinational corporation, this problem posed itself as meeting the demands to achieve global scale against the needs of national markets and governments. Given different national principles of organizing the activities of multinational corporations, there was not a rapid convergence to a global structure. The growth of multinational corporations, especially in Europe, was stymied by these traditions. Continuing national differences and the failure to develop harmonized European corporate law deterred the emergence of pan-European multinational corporations. Many of the most important cross-European mergers in the 1970s failed within a decade. Instead, American multinational corporations, less bound by these separate national traditions, were able to create integrated European strategies.

Nevertheless, it was in Europe where new organizational structures developed to resolve the conflict between integration and differentiation. Multinational corporations from smaller countries offered the laboratory experiments that influenced the evolution of multinational corporations from larger nations. Because of the small size of many European countries, large firms in Sweden, Switzerland, and the Netherlands quickly became multinational corporations, and these corporations had far more assets and employees outside their home countries than within. Over the course of the last two decades of the twentieth century, organizational structures developed, labeled as transnational corporations, that distributed global responsibilities to national subsidiaries, created global teams and projects, and encouraged the transfer of ‘best practices’ across borders.

3. Diffusion of Organizational Knowledge

In its evolution the multinational corporation is not without serious contradictions. Evolving from its national context, the multinational corporation employs large numbers of employees of diverse nationalities and ethnicities. Westney (1993) notes that a subsidiary is, thus, caught between the institutional pressures to conform to the company norms and values, as well as to the cultural and social influences of its local national environment. At the heart of the evolution of the multination corporation, thus, lies the tension between national institutions and the fragile emergence of a global culture.

The international evolution of the organizational structures of American multinationals mirrored, as we noted above, the broader diffusion of organizational technologies in the home market. The initial investments by a firm took place, often, on the basis of opportunity and the extension to familiar countries. Much like the ethnic trading communities dating back to the earliest times, the inexperienced multinational corporation preferred countries that are culturally similar to what their managers know at home (Johanson and Vahlne 1978). In these countries, they often established foreign enclaves where their expatriate managers could live in the simulated familiarity of their home environments.

The relevance of the second definition of foreign direct investment as the transfer of organizational knowledge is critical to understanding the powerful conflicts posed by the multinational corporation. The multinational corporation, competing often on superior technologies and managerial capabilities, serves
as a conduit of knowledge across borders. Some kinds of technologies can be purchased on markets and the flow of licensing payments attached to the sale of the right to use a technology constitutes a nontrivial flow in international balance of payments. Technology can also flow by the movement of people. Just as English craftsmen were imported by continental Europe in the early industrial period, there exists an international market for managers and skilled workers.

However, some knowledge is embedded in organizations and can only be transferred as an organization across borders. The issue is more than whether knowledge is tacit or explicit, for this pertains to the knowledge held by individuals too. The central feature of the diffusion of knowledge by the multinational is that it transfers the knowledge of how to organize and how to coordinate people and across divisions. These organizing principles then provide the capabilities for the firm to achieve quality, speed products to the market, or lower costs (Kogut 1992). For this the multinational corporation is required, which through its own activities or in joint ventures with host companies, transfers the organizational knowledge.

The global market for knowledge extends also to consultants. The large American consulting practices often have their origins in their acquisition of American organizing principles that they then transferred around the world. The international diffusion of the divisional structure, for example, reflects the knowledge acquired in American consulting practices that was then sold in abroad. Channon (1973), for example, observes that half of the firms he observed as adopting the divisional structure relied upon the consulting services of the same American firm. Similarly, British and then American banks spread throughout the world on servicing the needs of their expanding home clients, and then on transferring their practices into these countries (Jones 1993).

It is important to realize that once this transition period passed, the multinational corporation and multinational service firms, such as those engaged in consulting, banking, and advertising, had permanently and historically changed the global economy. No longer was the flow from the US to Europe, or from one ethnic community to another. The network of connections developed by the multinational corporation permitted the flow of ideas and practices among and between countries. Thus, whereas it took half a century for American practices to diffuse to Europe, the introduction of quality circles from Japan happened within a decade or two of their original innovation. In Europe, Japanese innovations were often diffused by American firms, including consulting companies.

However, this image of the free flow of knowledge needs to be strongly conditioned on the continuing importance of national institutions. Nations consist of defined cultures and economic and social institutions, such as unions, financial systems, and religious values. These institutions interact and their complex interactions causally influence behaviors. For example, the German system of centralized bargaining, enterprise clubs, strong banks, and social welfare has, until recent times, composed a national configuration of institutions that influences the capabilities of resident firms to be able to manufacture high quality goods for export markets (Streeck 1995, Soskice 1990). The introduction of Japanese and American teams confronts in Germany the presence of existing work councils. These councils are fairly rigid features of the German environment. Since Japanese methods may not be effective without teams as complementary factors in organization, this institutional refusal can effectively deter the diffusion of these methods. In short, national organizing principles are embedded in the wider social institutions. Inconsistency between these institutions and principles, then, is an empirical question of the bargaining strength of vested powers.

4. Economics and Politics of Multinational Corporations

Since the multinational corporation is definitional equivalent to foreign direct investment, theories of foreign direct investment must account for why one country invests in another and why this investment is carried out within organizational boundaries of a firm (see Buckley and Casson 1976, see Foreign Investment: Direct). In distinguishing between portfolio and direct investment, Hymer noted that firms operate at a disadvantage in foreign markets and hence they must have an offsetting competitive advantage to compete overseas. These advantages for overseas investments are the same ones that allow a firm to compete and grow in the home market. These observations have important implications. The first is that direct investment is the growth of the firm across borders and hence the firm expands internationally on what it has learned at home. This observation is the basis for the evolutionary theory of the firm. The second observation that Hymer made is that firms that expand overseas, because they have competitive resources, are also likely to be large and to belong to oligopolistic industries.

In these observations, we can understand the ambivalence expressed in popular and policy debates regarding the multinational corporation. Competition among multinational corporations often is the extension of their home domestic and oligopolistic rivalry that spills across national borders. In many global industries, the same company names dominate each country's list of the largest firms inside their national frontiers. No matter if it is Poland or France, Singapore or Mexico, the same multinational corporations will be found in the local oligopolistic industries (e.g., consumer goods or automobiles).
Because they are large even in their home markets, investments by multinational corporations can have a large impact on a host country (Caves 1974).

As a consequence, the multinational corporation often has been the subject of debates concerning national sovereignty and welfare. In recent decades, acquisitions have generally been the primary way by which multinationals invest in wealthy foreign countries, where the vast proportion of direct investment is concentrated. Given the size of a multinational corporation and occasional national importance of the targeted acquisition, even wealthy countries frequently evidence discomfort, if not outright public hostility, to multinational investments. Moreover, multinational corporations are sometimes the vehicles for foreign policies of their home or host country. The decision, for example, of the US to embargo technology and investment flows to Cuba, the former Soviet Union, Iran, and other countries periodically has caused conflict with other countries.

Multinational corporations are especially problematic in developing countries. By definition, developing countries are relatively poor, thus both in need of capital and yet concerned over their loss of independence. As discussed above, the history of multinational corporations in developing countries is marked by its origins in policies of imperialism and colonialism. Especially in Latin America, where a school of thought labeled Dependencia has been influential, the concern over dependence on the United States resulted in efforts to curb the power of the multinational corporations by restricting the amount of equity ownership a foreign firm could hold in a domestic company or by prohibiting investment in certain sectors. Mexico’s constitution forbids foreign investment in the oil industry; Brazil pursued for a long time a policy to restrict foreign participation in the electronics industry.

The other side of the coin is that multinational corporations bring investment and technology to the foreign country. Vernon (1966) hypothesized that innovations start in wealthy countries. As the market is saturated and as oligopolistic rivalry increases, multinational corporations are pushed out from their home markets to expand abroad in new markets and to locate less expensive places. Thus, Vernon seized both sides of the debate, recognizing the value of the transfer of technology but also emphasizing the oligopolistic nature of multinational investment.

It is, in fact, difficult to draw simple conclusions regarding the relationship of foreign investment and national growth. Countries such as Singapore, Malaysia, and Thailand have encouraged foreign direct investment actively. The growth in China’s coastal sector is indisputably linked to the massive investments by multinational corporations. However, historically Japan and Korea have pursued more cautious policies regarding investments by multinational corporations. In these countries, the state has often negotiated the terms for entry by multinational corporations, sometimes requiring licensing to domestic competitors as a price. The efficacy of such policies for these countries is much disputed. However, for many other countries, the intervention of the government in demanding licenses unquestionably leads to internal corruption and to insufficient domestic competition.

There are many channels by which a country can absorb foreign technology and managerial techniques. Most of the evidence shows, however, that prohibitions on the in-flows of direct investment can be very costly for many countries. With their domestic industries still to be developed, a developing country requires substantial investment. Some countries, primarily in Asia, have been able to achieve very high savings rates to finance their industries without direct investment. Moreover, high savings rates, plus political stability, create growth, and growth attracts foreign portfolio capital. A poor country that prohibits foreign direct investment but does not have high rates of saving is entirely dependent upon portfolio capital. The history of debt and currency crises in the 1990s convinced many poor countries that foreign direct investment was a preferable means of attracting capital, because it could not be easily pulled out of a country on short-notice in response to a financial crisis.

However, multinational corporations also respond to the volatility in the global market. This volatility derives from changes in exchange rates, politics, and productivity. Once having achieved sufficient experience and having established subsidiaries around the world, the multinational corporation might choose to close a plant in one location and open plants in new locations. Of course, such actions might provoke a response by labor, but historically, labor has been organized by national, not by international, organizations (Martinelli 1975). Yet, there is also the possibility that locations lose some kinds of plants but gain more sophisticated investments. Cantwell (1999) proposed that some regions and countries pull multinational investments. Yet, it has long been noticed that foreign direct investment among developed countries flows to high cost locations. Regions such as Silicon Valley, Baden-Wuertemberg, and Singapore attract multinational investments not because wages are low, but because productivity levels are high and workers are well trained. In many cases, developing countries have given rise to their own multinational corporations acting in the region and sometimes globally (Lall 1983). In this sense, the multinational corporation acts as a training center in the developmental strategies of emerging economies.

5. Globalization

The peculiar conflict in the world economy is the growing trend toward economic integration without a
concomitant growth in global political and social institutions to regulate and arbitrate this trend. Multinational corporations are rapidly adopting advanced information technologies to increase the efficiency and capabilities of their operations. Valuable information, such as software or financial services, is transmitted digitally. Governments are often unable to tax the value of these services or control their content. Moreover, information technologies, often supported by private telecommunication networks, permit, for example, a unit in Germany to control the manufacturing operations located in Brazil.

Digital technologies permit a high degree of integration and, in the short run, aggravate the gap between countries that are rich and those that are poor and do not have the infrastructural capabilities. However, these technologies are beginning to have profound effects on some regional economies. In parts of China, India, Israel, and elsewhere, advanced satellite transmission transmits digitally encoded work between their sites and other organizational units of multinational corporations located elsewhere. Whereas before the Indian engineer from Bangalore might have tried to migrate to the US to bring his human capital to a more attractive labor market, the increasing wages and job prospects in India are encouraging many to stay at home and participate digitally in the world economy.

These trends have unclear effects on multinational corporations. They permit more easily the development of projects that are in continual development, as work passes from one unit to the next as the day advances. They also allow more easily the coupling of less expensive labor in one country with more expensive skilled labor in another.

At the same time, the origins of the multinational corporation laid in its ability to organize labor on a worldwide basis on principles other than ethnic and cultural identities. This organization has never been without cost and risk. The growth of a world digital economy permits alternative ways by which labor can cooperate and be coordinated on a world basis. The intriguing question at this point in history is whether the multinational corporation, though still a vital presence in the world economy, nevertheless will recede relatively in importance as information technologies reduces the meaning of geographic distance.

See also: Development and the State; Diaspora; Globalization: Geographical Aspects; Globalization: Legal Aspects; Globalization: Political Aspects; Globalization, Subsuming Pluralism, Transnational Organizations, Diaspora, and Postmodernity; International Business; International Law and Treaties; International Marketing; International Organization; International Trade: Economic Integration

Bibliography


Channon D F 1973 The Strategy and Structure of British Enterprises. Harvard University, Boston


Lall S 1983 The New Multinationals: The Spread of Third World Enterprise. Wiley, Chichester, UK


10203
Multiple Imputation

Missing data are ubiquitous in social science research. For example, they occur in a survey on people’s attitudes when some people do not respond to all of the survey questions, or in a longitudinal survey when some people drop out. A common technique for handling missing data is to impute, that is, fill in, a value for each missing datum. This results in a completed data set, so that standard methods that have been developed for analyzing complete data can be applied immediately (see Nonsampling Errors and Statistical Data, Missing).

Imputation has other advantages in the context of the production of a data set for general use, such as a public-use file. For example, the data producer can use specialized knowledge about the reasons for missing data, including confidential information that cannot be released to the public, to create the imputations; and imputation by the data producer fixes the missing data problem in the same way for all users, so that consistency of analyses across users is ensured. When the missing-data problem is left to the user, the knowledge of the data producer can fail to be incorporated, analyses are not typically consistent across users, and all users expend resources addressing the missing-data problem.

Although imputing just one value for each missing datum satisfies critical data-processing objectives and can incorporate knowledge from the data producer, it fails to achieve statistical validity for the resulting inferences based on the completed data. Specifically, for validity, the resulting estimates based on the data completed by imputation should be approximately unbiased for their population estimands, confidence intervals should attain at least their nominal coverages, and tests of null hypotheses should not reject true null hypotheses more frequently than their nominal significance levels. But a single imputed value cannot reflect any of the uncertainty about the true underlying value, and so analyses that treat imputed values just like observed values systematically underestimate uncertainty. Thus, using standard complete-data techniques will result in standard error estimates that are too small, confidence intervals that undercover, and P-values that are too significant; even if the modeling for imputation is carried out carefully. For example, large-sample results by Rubin and Schenker (1986) show that for simple situations with 30 percent of the data missing, single imputation under the correct model followed by the standard complete-data analysis results in nominal 90 percent confidence intervals having actual coverages below 80 percent. The inaccuracy of nominal levels is even more extreme in multiparameter problems (Rubin 1987, Chap. 4), where nominal 5 percent tests can easily have rejection rates of 50 percent or more when the null hypothesis is true.

Multiple imputation (Rubin 1987) retains the advantages of single imputation while allowing the data analyst to obtain valid assessments of uncertainty. The key idea is to impute two or more times for the missing data, using independent draws of the missing values from a distribution that is appropriate under the posited assumptions about the data and the mechanism that creates missing data, resulting in two or more completed data sets, each of which is analyzed using the same standard complete-data method. The analyses are then combined in a simple generic way that reflects the extra uncertainty due to having imputed rather than actual data. Multiple imputations can also be created under several different models to display sensitivity to the choice of missing-data model.

1. Theoretical Motivation for Multiple Imputation

The theoretical motivation for multiple imputation is Bayesian (see Bayesian Statistics), although the procedure has excellent properties from a frequentist perspective. Formally, let Q be the population quantity of interest, and suppose the data can be partitioned into the observed values X_{obs} and the missing values X_{mis}. If X_{mis} had been observed, inferences for Q would have been based on the complete-data posterior density p(Q|X_{obs}, X_{mis}). Because X_{mis} is not observed, inferences are based on the actual posterior density p(Q|X_{obs}), which can be expressed as

\[ p(Q | X_{obs}) = \int p(Q | X_{obs}, X_{mis}) p(X_{mis} | X_{obs}) dX_{mis} \]  (1)

Equation (1) shows that the actual posterior density of Q can be obtained by averaging the complete-data posterior density over the posterior predictive distribution of X_{mis}. In principle, the set of multiple imputations under one model are repeated independent draws from p(X_{mis} | X_{obs}). Thus, multiple imputation allows the data analyst to approximate Eqn. (1) by separately analyzing each data set completed by imputation and then combining the results of the separate analyses.