The Historical Fertility Transition: A Guide for Economists

TIMOTHY W. GUINNANE

The historical fertility transition is the process by which much of Europe and North America went from high to low fertility in the nineteenth and early twentieth centuries. This transformation is central to recent accounts of long-run economic growth. Prior to the transition, women bore as many as eight children each, and the elasticity of fertility with respect to incomes was positive. Today, many women have no children at all, and the elasticity of fertility with respect to incomes is zero or even negative. This paper discusses the large literature on the historical fertility transition, focusing on what we do and do not know about the process. I stress some possible misunderstandings of the demographic literature, and discuss an agenda for future work. (JEL I12, J13, N30)

1. Introduction

Prior to the historical fertility transition, which took place in most of Europe and North America between the late eighteenth and early twentieth centuries, married women could expect eight or more births. Today many women do not have children at all. For centuries, the elasticity of fertility with respect to incomes was positive. Today it is zero or even negative. The fertility transition forms a major part of the process that brought the European, and later other, economies from slow to rapid and sustained growth. This article discusses the empirical literature on the historical fertility transition: what we know and what we still need to learn.

Current interest in the broad area of demographic behavior and long-run growth reflects several intellectual influences, including Robert E. Lucas’s lectures on growth (Lucas 2002). Early efforts in this area usually

* Yale University. My research on this and related topics has been generously funded by the National Institutes of Health (U.S.), the Leverhulme Trust (U.K.), and the Economic and Social Research Council (U.K.). This paper has been presented in seminars at Brown University, the University of Colorado (Boulder), the University of Santiago de Compostela, and the Autonomous University of Barcelona. I have learned much about these issues from working with and talking to John Brown, Christoph Buchheim, Oded Galor, Carolyn Moehling, Thomas Mroz, Sheila Ogilvie, Barbara Okun, Cormac Ó Gráda, Jean-Laurent Rosenthal, Mark Rosenzweig, Paul Schultz, William Sundstrom, Nathan Sussman, James Trussell, Christopher Udry, and David R. Weir. For comments and suggestions on this paper, I thank the editor and three referees, as well as Robert Allen, Stanley Engerman, James Fenske, David Miteh, Tommy Murphy, and Marianne Wanamaker. My interest in this subject is due to Paul David, and any (good) ideas here probably reflect his influence more than is reflected in the citations.
embed the microeconomic model of fertility decisions developed by Gary S. Becker in a framework that allows feedbacks from the economy to fertility decisions. Perhaps the leading paper in this area is Becker, Kevin M. Murphy, and Robert Tamura (1990). A more recent body of research focuses explicitly on how economies transition from a “Malthusian” economy of high fertility and little growth in per capita incomes to one in which fertility is much lower and per capita incomes grow rapidly. This “Unified Growth Theory” (hereafter UGT), developed by Oded Galor, David N. Weil, and others, integrates a microeconomic model of the demand for children with feedbacks from a model of growth. Other, related literature either considers fertility and growth in a different way, or examines related questions such as the role of children in the labor market.

These discussions have provoked renewed interest in the empirics of the historical fertility transition: when and why it took place, and how rapidly it created the low fertility we now see in most wealthy countries. Growth theorists are to be commended for trying to explain this historical fact and for highlighting the importance of issues that were previously of interest primarily to economic historians and to others, such as demographers, outside the economics profession. Unfortunately, some recent discussions misconstrue the demographic literature. The use of certain theoretical ideas has also led some to believe that the relevant propositions have actually been tested in historical contexts, when that is unfortunately often not the case. Fully understanding the fertility transition will require much more empirical research. This paper has two goals. I provide an overview of the historical fertility transition, and then discuss the main hypotheses that are current in the economics and economic history literature. Throughout, I set aside three other issues. First, most current economic research stresses the implications of demographic change for economic growth and development. That issue lies beyond the scope of this paper. Second, economists typically think of fertility decline as the reduction in the number of children born to a woman or to a couple. Demographers and others stress heterogeneity in the way fertility declines, for example, whether couples reduce the number of surviving offspring by spacing their childbearing or ending childbearing before that is biologically necessary. I discuss this issue only where important to understanding the evidence on the fertility transition. Finally, the fertility transition in developing countries since World War II has been studied far more intensively than its historical counterpart. This paper focuses on the earlier, historical episode, which is most relevant to what theorists have in mind when modeling the Industrial Revolution.

2. The Basic Contours of the Historical Fertility Transition

Figure 1 reports fertility experience for the period 1800–1970 for five major countries: France, England and Wales, Germany, the United States, and Italy. The measure reported in figure 1 is the Crude Birth Rate (CBR), defined as the number of births per thousand per annum. This paper focuses on the first four countries; Italy is included in figure 1 only to suggest the heterogeneity of historical experience. Figure 2 focuses on the single case of Germany to highlight the relationship between fertility and mortality decline. Ignoring the heterogeneity in figure 1 for the moment, we see fertility declining starting in the

\[ \text{Deaths in } \text{Figure 2} \text{ are reported as the Crude Death Rate (CDR), defined analogously to CBR. The Crude Rate of Natural Increase (CRNI) is defined as CBR – CDR.} \]
eighteenth or nineteenth century, and this decline accelerates in the second half of the nineteenth century. The two world wars produced dramatic, temporary reductions in fertility, and the post–World War II period saw a “Baby Boom.” By the 1970s, most of Western Europe experienced the emergence of “low–low” fertility. Today, some OECD countries have fertility rates too low to sustain population growth through natural increase alone.² The relationship between fertility and mortality declines differs across these countries, but the German experience

² Hans-Peter Kohler, Francesco C. Billari, and José Antonio Ortega (2002) discuss the emergence of very low fertility in the 1990s; Joshua R. Goldstein, Tomáš Sobotka, and Aiva Jasilioniene (2009) discuss a recent, partial reversal.
detailed in figure 2 is fairly typical. For most of the nineteenth century, birth rates exceeded death rates and population grew (even with, as in the German case, extensive emigration). This is no longer true; for much of the late twentieth century, German rates of natural increase were negative.

Most countries outside Europe and North America did not experience fertility declines until after World War II. Timing within Europe and North America is less clear. Some economists have accepted the view that the fertility transition took place all at once across Europe. Most economists would not admire the data and methods upon which that assertion is based. The all-at-once view has important implications for causality. Some scholars invoke the claim of simultaneous fertility transitions to support their view that economics has little to do with the fertility transition: “Clearly the simultaneity and speed of the European transition makes it highly doubtful that

Figure 2. Fertility and Mortality in Germany
(Number of events per thousand population)

Source: Crude birth rates as reported in Mitchell (1980).

3 Galor (2005a, footnote 33) accepts this view, for example. Galor is citing the results of a large project undertaken at Princeton University in the 1960s and 1970s. The Princeton conclusion reflects serious problems with sources, measures, and econometrics. The Princeton project devised a series of indices of overall fertility, marital fertility, illegitimate fertility, and the contribution of marriage patterns to fertility. The index of marital fertility, $I_m$, is scaled such that it would equal one if the population in question had fertility equal to the very high level of the Hutterites in the 1920s. See appendix B to Ansley J. Coale and Susan Cotts Watkins (1986) for definitions. $I_c$ does not perform as claimed in Monte Carlo studies (Timothy W. Guinnane, Barbara S. Okun, and James Trussell 1994). This index’s performance is crucial to the claim that fertility started to decline everywhere at once. John C. Brown and Guinnane (2007) critique the statistical exercises reported in that project’s publications.

4 Thus it is puzzling to find an economic historian claiming, counter to the evidence, that “...the timing of the demographic transition in Europe and the United States places it circa 1890” (Gregory Clark 2007, p. 225).
any economic force could be found which was powerful enough to offer a reasonable explanation” (John Cleland and Christopher Wilson 1987, p. 18).

Figures 1 and 2 report national aggregates. Several studies document the existence of fertility control among small groups as early as the seventeenth century. These “forerunners” were usually urban elites or members of minority groups such as Jews (Massimo Livi-Bacci 1986). More generally, research based on either subnational aggregates or micro data often find earlier fertility declines than in national data. The Princeton studies report earlier fertility declines in cities, for example, but stress the behavior of national populations. John E. Knodel (1988) reports that couples married 1800–24 in five of the fourteen German villages he studied exhibited significant control of marital fertility.

There are more general reasons for skepticism about simultaneous fertility transitions. Most scholars agree that France’s fertility decline started in the early nineteenth century at the latest. David R. Weir’s (1994, table B3) careful construction of the Princeton indices shows that the decline in French fertility began in the late eighteenth century. In the early nineteenth century, U.S. birthrates were higher than elsewhere, but falling rapidly. By the end of the nineteenth century, U.S. fertility fell below that of much of Western Europe. France and the United States are not esoteric exceptions one can ignore.

Another concern reflects the CBR so often relied upon in economics research. In many circumstances, it is the only measure available or, as in my figure 1, the only measure available that is defined consistently across a set of countries for the right period. Yet it has serious weaknesses, especially applied to this period of rapid economic and demographic change. The CBR does not account for age-structure or marriage patterns, both of which changed significantly in some European populations during this period. Consider England and Wales, where from 1871 to 1911 there was a disproportionate increase in the number of women of child-bearing age. The overall population increased by about 60 percent, while the female population aged 15–44 increased by 77 percent. The shift in age-structure in the absence of a fertility reduction would have increased the CBR by 11 percent. By failing to take into account the changing population age-structure, the CBR understates the change in behavior. Some European societies experienced significant changes in marriage patterns during this period that make the CBR a poor guide to marital fertility. For example, between 1871 and 1911, the proportion of English women who had never married by ages 25–29 rose from 0.37 to 0.44 (John Hajnal 1953, table 3). Mean age at marriage for English males rose over the same period by 1.2 years (to 27.6), while at the same time in Germany it fell by nearly 1 year (to 27.9). The changes for women are similar but slightly less pronounced (Josef Ehmer 1991, table 1).

6 This precise effect is not present in every population represented in figure 1. This is the point. The figures in question come from the decennial censuses of England and Wales for 1871 and 1911, and the Registrar-General’s reports on births, deaths, and marriages. I have taken them from B. R. Mitchell (1980, series B1, B2, B5, and B6).

7 Cross-sectional differences in marriage patterns at the time were especially large. In 1900, England and Ireland had about the same CBR, but the latter was achieved by relatively few couples having very large families (Guinnane 1997).

---

5 J. David Hacker (2003) questions whether the United States really had an early fertility decline. He is right to stress the weakness of the sources for the early nineteenth century, but the methods he prefers are fragile in their own way, and even he suggests a U.S. fertility transition well before the late nineteenth century.
Figure 3 reports a better measure, the Cohort Fertility Rate (CFR) for the five countries in Figure 1. CFR is the average number of children born to women in a given cohort, and thus requires age-specific fertility for a cohort’s entire reproductive life. Its cross-section analogue, the Total Fertility Rate (TFR), is the sum of current age-specific fertilities, and thus reports the number of children born to a woman who experiences current age-specific fertility throughout her life. Reliable data needed to compute the CFR do not go back very far into the nineteenth century, and are not available even then for most countries. But Figure 3 suffices to suggest that the experience of specific cohorts born in the nineteenth century is not well-captured by CBR.8 What looks like constant fertility (as measured by the CBR) may well reflect two offsetting trends. Figure 4 uses the exceptional French data to illustrate just this point: France’s precocious marital fertility decline was partially offset by a mid-century marriage boom.

8 TFR estimates for the United States start at 7.04 in 1800 and fall to 3.56 by 1900. These figures, just like the CBR in Figure 1, are for the white population only (Michael R. Haines 2000a, table 4.3).
3. From Malthus to Becker

Most of the literature uses Malthusian models to understand the relationship between population and the economy before the fertility transition, and views the fertility transition as an escape from the Malthusian world. Figure 5 shows that escape: until the early nineteenth century, there was a positive relationship between fertility and the real wage in England and Wales. Then the relationship breaks down: higher wages, the product of capital accumulation and technological change, no longer translated into higher fertility.

Two distinct approaches to the Malthusian world exist in the literature. An older version stresses Malthus’s own argument, that is, that the regulation of births reflected the regulation of marriage. In this model, marital fertility depends only on age at marriage and the proportions who marry. The lifetime fertility of any given woman is a stochastic function of her age at marriage.
Central to this version of the Malthusian model is the “European marriage pattern” in which young adults deferred marriage until well after puberty, often into their middle and late twenties, and as much as ten to twenty percent of some cohorts never married at all. Couples could not marry until they could support themselves and their offspring, implying that marriage decisions depended on the real wage in young adulthood.\footnote{Most studies find a substantial proportion of pregnant brides, but children actually born outside of marriage rarely accounted for more than 5 percent of all births. For the European marriage pattern, see Hajnal (1965, 1982). Most scholars accept Hajnal’s European Marriage Pattern as a stylized account, but it is not clear which parts of Europe it describes. E. A. Wrigley and R. S. Schofield’s (1981) book is an extended argument that England’s demographic system functioned in a particularly benign way because of the strength of the relationship between the real wage and marriage patterns. Earlier studies had stressed mortality changes as the driving force in English population history. Malthusian models in Chinese population history are more controversial, partly because Chinese marriage patterns appear to have been different. James Z. Lee and Wang Feng (1999) provide an overview of the debate and references to the literature.} In this version of the Malthusian model, the fertility transition reflects a shift from controlling marriage to controlling fertility within marriage. More recently, many economists (and much of the recent growth theory) set aside the marriage issue and thus model fertility without concern for its underlying determinants.\footnote{This vision underlies Galor and Weil (2000), as well as most other UGT models.}

3.1 The Demand for Children

Virtually all economic analysis of fertility today starts from Becker’s model of the demand for children.\footnote{The age-structure change in England noted above is one such echo effect. Ronald D. Lee and Michael Anderson (2002) rely on a state-space representation to contend with the integration and endogeneity problems implied by the model. This paper presents a clear overview of earlier literature as well as the modeling problems at issue. More recently, Esteban A. Nicolini (2007) relies on a simple VAR model that he identifies by an assumption of “contemporaneous stickiness.” Niels Framroze Møller and Paul Sharp (2008) are unusual in the recent literature in explicitly modeling marriage rates. There are also data reasons to prefer the short-run model: often for historical situations we know the number of events (births, deaths, marriages) but not the population size, and thus cannot compute demographic rates. Lee (1981, 1985) explains the logic of the short-run models. Weir (1984) used this approach to challenge Wrigley and Schofield’s interpretation of their English evidence. Guinnane and Sheilagh Ogilvie (2008) apply this approach to some German villages for the period 1634–1870 and provide references to other efforts of this sort.} Becker’s insight was to analyze the demand for children using the tools of consumer choice. The model yields important insights for the fertility transition. Observers have long noted that fertility tends to be negatively correlated with income in the cross-section, and, since the beginnings of the fertility transition, over time. Becker’s model implies that this is a standard substitution effect, that children are not inferior goods: wealthier couples have higher opportunity costs of time, and time is a major cost of child-rearing. The simple version of Becker’s model starts with a household utility function $U = U(n, Z)$, where $n$ is the number of children and $Z$ is a vector of all other commodities. The household maximizes this utility subject to a standard budget constraint. Increases in child costs induce substitution

\begin{equation} \end{equation}
away from children and toward the $Z$s. A pure increase in income raises the number of children demanded, as we expect. But if that higher income reflects rising wages, then that increased wage may show up as an opportunity cost of having children, and reduce the number of children demanded via the substitution effect.\footnote{In a number of articles often called collectively the “Easterlin synthesis,” Richard A. Easterlin integrated Becker’s approach with a better appreciation of the costs of fertility control as well as the biological limits on reproduction. See especially Easterlin (1978).}

Later interest in Becker’s model focuses on the possible trade-off between the \textit{number} of children and their \textit{quality}, usually called the Quantity/Quality or “Q–Q” model. (Child quality is usually taken to mean child’s health or education.) The Q–Q model starts from a household utility function of the form $U(n, q, Z)$, where $q$ is the quality of each child. Becker (1981, pp. 107–08) divides child costs into three categories. Some costs depend only on the \textit{number} of children: an example of this, $p_n$, would be the costs associated with the mother’s pregnancy and delivery.\footnote{This discussion follows Becker (1981) and uses his notation.} Another cost is related to child

---

\textbf{Figure 5. Crude Birth Rates and Real Wages}

Notes: This figure differs from Wrigley and Schofield (1981, figure 10.1) in two ways. I plot the CBR, not the Gross Rate of Reproduction. The real wage index here is Robert Allen’s “labourers” index, rather than the Phelps Brown-Hopkins index. The series plotted are centered eleven-year moving averages. Allen’s index can be found at: http://www.economics.ox.ac.uk/members/robert.allen/WagesPrices.htm.

Source: Based on Wrigley and Schofield (1981), figure 10.6.
quality, but does not depend on the number of children, as it goes to purchase household public goods: examples of $p_q$ include books that children could share. A final cost, $p_c$, is the cost of augmenting the quality of any child. The household’s budget constraint is then:

$$p_q q + p_c nq + \pi_z Z = I,$$

where $I$ is household income and $\pi_z$ is the price of the $Z$s. The marginal rate of substitution between quantity and quality now depends on the ratios of fixed to variable costs for quantity and quality respectively, as well as on the ratio of marginal variable cost to the average variable cost of quality. The substitution effects between quantity and quality are stronger in the Q–Q model than in the original Becker model. Consider an increase in $p_n$. The household will substitute away from numbers to both child quality and $Z$s, as one would expect. But because of that interaction term $p_c nq$, the shadow cost of $n$ depends on $q$, so a reduction in child numbers raises the shadow cost of numbers even more, inducing more substitution of quality for quantity.

This Q–Q model has considerable appeal in historical circumstances. We observe sharp fertility declines that seem to reflect small changes in the economic environment. For example, as Becker noted himself, within this model a modest reduction in the cost of contraception could induce a shift from $n$ to $q$, which seems consistent with the historical evidence. Several recent papers use historical data to test the Q–Q model. Sascha O. Becker, Francesco Cinnirella, and Ludger Woessmann (2009) use toughening of compulsory schooling laws in Prussia in the late 1840s to study the fertility effects of a reduction in the price of child quality. In their district-level data, areas with higher enrollment rates had lower fertility. This finding seems consistent with the central trade-off in the Q–Q model. Hoyt Bleakley and Fabian Lange (2009) take a different approach. They examine a program that largely eradicated intestinal worms among children in the early twentieth-century U.S. South. They find that the reduction in the prevalence of worms, which they interpret as a shock to the cost of child quality, increased school enrollment rates and reduced fertility. This again is consistent with the Q–Q model.

4. **Explanations and Evidence**

We can group the many economic explanations offered for the historical fertility transition under six headings. The first is an exogenous decline in infant and child mortality. The second turns on innovations in the technology of contraception, or more widespread availability of contraceptive devices. The third looks for increases in the direct cost of childbearing. The fourth explanation is based on changes in the opportunity costs of child-bearing. The fifth looks for a net increase in returns to child quality. The sixth argument assumes that children were an important way to ensure against risk and to provide for old age, and that the rise of state social insurance as well as private insurance and savings vehicles led households to substitute out of children. I consider these explanations in ceteris paribus fashion.

4.1 "**Demographic Transition Theory**” and the Role of Mortality Decline

A long tradition assigned to mortality decline a causal role in the fertility transition.

15 The Q–Q model may get more attention than it warrants. Mark R. Rosenzweig and Kenneth I. Wolpin (1980) show that Becker’s original model (with $p, nq$ constrained to zero) generates nearly all of the testable implications that are identified with the Q–Q model. See also T. Paul Schultz (1981, pp.166–69).
Frank Notestein (1945) argued that couples in high-mortality societies have a lot of births to ensure a surviving brood of the desired size. An exogenous mortality decline induces couples to have fewer children because they do not need so many “spares.” Notestein’s account was motivated by the experience of developing countries after World War II, where public-health interventions first reduced infant and child mortality, and in some places those developments were followed by declines in fertility. Most European countries also experienced a significant mortality decline in the nineteenth century. Historians and others still debate the causes of the historical mortality decline, but most scholars stress some combination of better food supplies, improvements in public health systems (such as clean water supplies and food-safety measures), and modest results from medical interventions (such as vaccines against smallpox). Some of these developments reflect local decisions about public-good investments, but it is plausible to view them as largely exogenous to any couple’s decision making.

This mortality decline was concentrated in the early years of life. A woman born in the United States in 1850 had an expectation of life of 39.4 years. A five-year-old girl in that year had an expectation of life of 50.8 years, and a twenty year-old could expect to live an additional 39.8 years. In 1910, these figures were nearly the same for those who had already survived the dangerous early years. A newborn girl could expect to live 54.7 years, and a five-year-old girl, 57.4 years. A twenty year-old woman could expect a further 40.7 years, not even a full year more than in 1850.

The Notestein argument does not fit the timing of the historical declines in fertility and mortality. Fertility in the United States declined for decades before any noticeable decline in mortality. The TFRs reported by Haines (2000b, table 8.2) decline from the early nineteenth century; there is no sustained fall in the infant mortality rate, on the other hand, until the 1890s. French experience was similar, with a fertility transition preceding mortality declines. In other places, such as Germany (figure 2), the fertility and mortality declines took place at roughly the same time. This does not rule out a role for exogenous changes in mortality as a causal force, of course, but it suggests that Notestein-style account explains only part of the change. The total fertility rate in the United States in the early nineteenth century was about seven. Even if thirty percent of children then died in infancy or childhood, this implies that households wanted a surviving brood of four or five. By the end of the nineteenth century, in contrast, white, urban women in the United States were increasingly having just two children (Paul A. David and Warren C. Sanderson 1987).

“Demographic transition theory” also has two theoretical flaws. Part of the decline in infant and child mortality is endogenous to the fertility decline. There are several lines of argument here, all of which assume that parents can assert some influence on their children’s mortality risks by providing health-enhancing resources. In a historical context, these resources include breast-feeding (which costs mother’s time, but isolates an infant from possibly contaminated water and

---

16 Thomas McKeown (1976) argued that prior to about 1900, medical science had done little to increase human longevity, and concluded that the observed mortality declines to that point reflected direct and indirect effects of better nutrition brought about by higher incomes and better food supplies. Robert William Fogel (2004)’s more nuanced account also stresses the role of nutrition. For an introduction and overview to this issue see, Angus Deaton’s review of Fogel’s book in the Journal of Economic Literature, 44(1): 106–14.

17 Historical Statistics of the United States series Ab657, 659, and 661. Mortality figures for the United States are complicated by the lack of complete death registration statistics, but the basic patterns noted in the text are robust.
food supplies), other nutrition, and protection from danger such as hearth fires. In the Q–Q model, reduced infant and child mortality could also reflect changes in other costs. For example, improved contraceptive technology (discussed next) could allow parents to more tightly control the link between actual and desired fertility. Parents might have a smaller number of children and care for them more intensively, in effect not relying on high mortality to cull their brood to its desired size.

The second theoretical problem with the “demographic transition” account is that even a fully exogenous reduction in infant mortality would have two, countervailing effects. An exogenous mortality decline reduces $p_n$, and thus makes child numbers cheaper relative to both child quality and $Z_s$. An exogenous mortality decline could actually raise fertility net of child mortality.

4.2 Innovations in Contraceptive Methods

A second explanation for the fertility transition implies that couples long wanted smaller families, and improvements in contraceptive methods made that goal easier to achieve.

Assessing this explanation is frustrated by lack of direct evidence on contraceptive practice for most of the period. The best we can do is to provide a broad characterization of those technologies and the constraints they put on couples. Indirect evidence shows that until the second half of the nineteenth century, most fertility control relied on a combination of withdrawal (coitus interruptus) and abstinence from sexual relations. The first “modern” methods appeared in the second half of the nineteenth century. These techniques relied partly on advances in medical understanding, but also on the invention of vulcanized rubber (in 1844). Applied to the production of condoms (in 1855), this new industrial process allowed couples to replace expensive and relatively unreliable condoms made from natural materials. (Most studies, in fact, conclude that prior to vulcanization, condoms were intended to prevent the spread of disease, not pregnancy). Depending on the place, condoms were widely available in barber shops, drug stores and other retail outlets. Vulcanized rubber was also the basis for the introduction of the diaphragm and similar barrier methods in the later nineteenth century.

Thus we have a new set of technologies, which would make contraception easier and more reliable, introduced not long before we observe the actual fertility decline in some countries (but well after France and the

---

18 Historical research has not always taken this issue seriously. Papers that do instrument for infant mortality in econometric analysis generally find that it dramatically reduces the apparent impact of mortality in an OLS framework. See Patrick R. Galloway, Lee, and Eugene A. Hammel (1998a), which also surveys earlier literature; Brown and Guinnane (2002); Tommy E. Murphy (2010a).

19 If this claim seems extreme, consider the practice of wet-nursing, which was extensive in France into the late nineteenth century. In many cases, one woman would take several urban babies soon after their birth, and travel with them by train to the location where they would be cared for. Some babies would die en route, from cold or hunger, and others would die from neglect at their destination. The practice was widespread and not limited to urban middle and upper classes. Catherine Rollet (1982, table 1) estimates that about ten percent of all French newborns, and about thirty percent of those born in the Paris region, were sent to a wet nurse in the late nineteenth century. Anne Martin-Fugier (1978, pp. 26–27) quotes a thirty percent mortality rate for wet-nursed infants in the Paris region earlier in the century. This is roughly twice the infant mortality rate for France as a whole at the time.

20 Angus McLaren (1978, pp. 25–27) among others notes that frequent condemnations of withdrawal in the eighteenth century suggest that the practice was already used as a form of contraception. James Woycke (1988, p. 11) concludes that throughout the nineteenth century, “... it was coitus interruptus that remained the most common contraceptive practice.” Gigi Santow (1995) provides the best recent account of this issue. The nineteenth century also witnessed the spread of “marriage manuals,” a euphemism for guides to sexuality, sexual health, and contraception. These guides appeared as early as the eighteenth century, although their circulation was at first limited (McLaren 1978, pp. 26–30). The usefulness of such guides in limiting fertility of course depends on their information being more accurate than what couples already knew from other sources.
United States). Robert T. Michael and Robert J. Willis (1976) first integrated the costs of averting unwanted births into the micro-economic model. Their model assumes that couples can affect births, which is a random variable, using methods that imply various utility and money costs. The couple’s optimization takes into account both the costs of contraception and the utility costs of having “too few” or “too many” children. (Thus the Q–Q trade-off enters their thinking indirectly; a couple with too many children may, because of the budget constraint, be forced to choose a lower level of quality than preferred.) Any contraceptive method implies both fixed costs (which must be “paid” to use the method at all) and a marginal cost (which depends on the number of births averted). A couple that wanted to avert three of an expected eight births would be happier with a relatively high marginal cost approach than would a couple that wanted to avert all but two of eight expected births.

The United States and many European countries at first made concerted efforts to limit the spread of contraceptive knowledge and technologies. In the United States, the “Comstock Laws,” a collection of state and federal statutes, made it illegal to disseminate both marriage manuals and contraceptive devices such as condoms. Similar measures were enacted in England, Germany, and many other European countries. The focus and practical enforcement of such laws were uneven, and in some cases, celebrated court cases probably advanced public knowledge of contraceptive methods. The policies could also be self-contradictory. Germany’s Lex Heinze (1900), for example, made illegal any public display or advertisement of objects intended for “obscene” use. But retailers could still sell condoms and other devices. At the same time, two other important German institutions, the Army and the Sickness Funds, were doing their best to encourage the use of condoms to stop the spread of venereal disease.

The effects of these restrictions on the fertility transition are not really known. Demographers today tend to argue that the availability of contraceptives and contraceptive information is the most important barrier to fertility decline in developing countries. Economists are more skeptical, stressing the incentives to reduce family sizes. In any case, three sets of useful surveys confirm that throughout the nineteenth and early twentieth century, withdrawal and abstinence remained the primary approaches used by married couples. Since these technologies had been available, essentially, throughout human history, it is unlikely that the condom and similar new methods played a strong role in the fertility transition. Legal restrictions probably mattered less than cost. Rubber condoms were at first expensive. Brown (2009a, pp. 15–16) estimates that in the early

21 Martha J. Bailey (2010) is one of the few careful empirical studies of the Comstock Laws, but deals with a period, the 1960s and later, that is well after the U.S. fertility transition. She concludes (p. 122) that in 1965, without the bans in place, marital fertility in the affected states would have been eight percent lower.

22 This view underlies the literature on the so-called KAP-gap. In many surveys in developing countries, a significant number of women report that they do not want any more children but are not using contraception, or that they had more children than they actually wanted. The precise reasons for this are debated, but many demographers think this fact reflects lack of access to contraceptives. Charles F. Westoff (1988) is an early, somewhat skeptical view. See also John Bongaarts (1991). The publications of the Alan Guttmacher Institute, on the other hand, argue that there is considerable “unmet need for contraception” today. See, for example, “Facts on Satisfying the Need for Contraception in Developing Countries,” April 2010.

23 Three sets of surveys, some of doctors and some of women or couples directly, all conclude that the primary techniques used by married couples were withdrawal and abstinence. See Brown (2009a, table 1) for a summary of three German surveys from the early twentieth century; David and Sanderson (1986, pp. 317–28) discuss the Mosher survey of American women born in the 1850s and 1860s, as well as later U.S. surveys; Robert Jütte (2003, p. 220) discusses a survey of French doctors from the 1890s.
twenty-first century, a year's supply of condoms cost a Berlin worker the equivalent of ten to fifteen days of wages. Other barrier methods such as the diaphragm required the attention of a trained medical professional. Cost may account in part for Santow's (1993) finding that *coitus interruptus* remained widely used well into the twentieth century, even in countries where alternatives were available. David and Sanderson (1986) develop a model of a couple's lifetime fertility as a renewal process, and use it to derive estimates of the number of live births a couple would experience in a twenty-year marriage under various assumptions about coital frequency and contraceptive failure rates. Their baseline couple (no contraception) would have about nine births if they had sexual intercourse, on average, five times per 24-day cycle. If this couple used a method with a 12.5 percent failure rate, and failed to use it about 10 percent of the time, they would have only three births in twenty years. This "method" approximates what we know about the use of *coitus interruptus* in modern populations. Conscientious use of condoms would get the couple below one birth. The methods available even prior to the fertility transition were sufficient to produce voluntary reductions of the magnitude we observe in the nineteenth and early twenty-first centuries.

4.3 Increases in the Direct Costs of Children

In the Q–Q model, increases in $p_n$ induce substitution toward both child quality and the $Z$ goods directly, and toward quality through the interaction term in the budget constraint. One logical possibility to explain the fertility transition is that the direct costs of child-bearing changed in ways that induced couples to have smaller families. The problem is that most costs did not change, over the relevant period, in ways that would produce the observed fertility decline. Most households in this period devoted the bulk of their expenditure to food, clothing, and housing. The real price of clothing dropped dramatically following the technological innovations of the Industrial Revolution, many of which were in textiles. Food prices varied over time and place, and protective tariffs on agricultural goods could raise the price of food in one country above its counterpart in others. But in general, food prices declined, which at a crude level would imply a reduction in $p_n$.

---

24 The contraceptive pill used today dates from the 1950s, with its first widespread use taking place in the 1960s. Historical sources also refer to efforts to induce abortion. Given social and later legal views of this practice, we cannot hope to know how common abortion really was during the period of the transition.

25 Their model is based on Mindel C. Sheps and Jane A. Menken (1973) and assumes that couples follow the same strategy over their entire lifetime. Michael and Willis (1976, table 2) report a similar exercise using different parameter values. The unwary economist may fall into a trap created by the way most demographers think of fertility control. To most demographers, the term "family limitation" does not mean a reduction in family size, it means a reduction in family size achieved in a particular way. A couple seeking to have only $N$ children can adopt a "stopping" strategy (have $N$ children before initiating any effort to restrict fertility) or a "spacing" strategy (reduce the probability of a birth right from the start of marriage, and thus spread out births throughout the fertility years of the marriage). Most studies in historical demography assume that only stopping is legitimate evidence of fertility control; evidence of smaller families achieved via spacing is attributed to motivations unrelated to family size. The reasons for this assumption lie beyond the scope of this paper, but amount to a concern about identification (see Louis Henry 1961). Many historical demographers also view rudimentary fertility control techniques as unsuited to a spacing strategy; dynamic models of family-building, on the other hand, imply that risk-averse couples using unreliable contraceptive methods will prefer spacing to stopping (David, Thomas A. Mroz, and Kenneth W. Wachter 1987). Several empirical studies show that spacing was widespread in the early stages of the fertility transition (see Lee L. Bean, Geraldine P. Mineau, and Douglas L. Anderton 1990 and Mroz and Weir 1990). The development of more reliable fertility control techniques in the nineteenth century could well have caused stopping to replace spacing. Works relying on the demographer's definition would identify a shift to stopping, holding number of children constant, as the onset of the fertility transition.
The only significant increases in direct costs took place because of urbanization. Most European countries as well as the United States experienced rapid urbanization during the nineteenth century. About six percent of the U.S. population lived in an urban place in 1800; in 1900 that was nearly forty percent (Haines 2000a, table 4.2). England was already very urban in 1801 (34 percent), and became even more so over the nineteenth century. By 1911, 79 percent of the English population lived in an urban center (Robert Woods 1996, table 3). France started out the period less urban, and while its cities did grow, it remained less urban than England or Germany. Urbanization in Germany was especially rapid in the late nineteenth century. Germans living in places with fewer than 2,000 people fell from 64 to 40 percent of the population between 1871 and 1910 (Hans-Ulrich Wehler 1995, table 71). In urban areas, housing costs exceeded costs in rural areas, but of course the decision to live in a city was up to a couple. Most studies find that urban fertility was lower than rural fertility in the nineteenth century, although the precise causation has not been established. Knodel (1974, table 3.2) reports that marital fertility in Berlin in 1867–68 was about 87 percent of that in rural Prussia; by 1905–06, Berlin’s fertility was half of rural Prussia’s and Prussian cities overall had fertility about 75 percent of rural Prussia’s. Haines (1989, table 2) estimates TFR in the urban United States in 1905–10 at about 2.7, with rural nonfarm TFR at 4.0 and rural farm TFR at 6.0. Once the fertility transition began, fertility usually fell first in urban areas, with rural areas then catching up.26

A second type of direct costs underlies a literature that started with the U.S. fertility transition. Easterlin’s (1976) famous explanation for the fertility decline in rural America rests on the rising costs of farmland as an area was settled. Suppose a farm couple wanted to establish each child on a farm similar to their own. As the price of local farmland rose, parents either had to send their children further west, where land was cheaper, or had to have fewer children. Easterlin argues that parents preferred to have fewer children and be able to settle them locally. He dates the beginning of the decline in New York State to 1805 and even Iowa, much further west, to 1835.27 Later research focuses on Easterlin’s assumption that parents wanted to give each child a fixed bequest. William A. Sundstrom and David (1988), for example, motivate their regression analysis using a bargaining model that presumes that a primary motivation for child-rearing is support in old age. In equilibrium, children can drive a harder bargain with their parents if they can point to better, off-farm opportunities. Cross-sectional regressions for U.S. states in 1840 show that fertility is negatively correlated with measures of nonfarm labor-market opportunities. Once such proxies are introduced, land prices have no influence on fertility.28

Child labor raises another source of variation in direct costs. In many societies children offset some of the direct costs to their parents by working either in parental income-generation activities (such as a farm) or by working in the labor market. Any change in children’s earnings would clearly

26 Alan Sharlin (1986) surveys urban–rural differences in European fertility, using the Princeton project’s data.

27 His fertility measure is the child–woman ratio, or the number of children age 0–9 per thousand women 16–44. This measure is sensitive to in- and outmigration of both children and adults, as well as to variations in infant and child mortality.

28 Susan B. Carter, Roger L. Ransom, and Richard Sutch (2004) provide additional evidence on this debate. This type of argument illustrates the problems with defining child quality. One could argue that Easterlin’s explanation is one where parents reduce child numbers when the costs of quality increased.
alter the net costs to their parents. Two important trends affected children’s earnings opportunities during the period in question. Most accounts imply that industrialization at first increased income-earning opportunities for children, because new technologies did not require physical strength. By the 1830s, large minorities of English children were working. Clark Nardinelli (1990, table 4.2) reports that in most English counties, at least one-quarter of children aged 10–14 were reported in the workforce. Some parts of the textile sector depended heavily on children. One Parliamentary inquiry reported that in cotton textiles, half of all workers were under 18, and 6.8 percent were under 10 (Nardinelli 1990, table 5.3). Wehler (1996, p. 254) notes that in German factories in this period, children could be fifteen percent of the workforce. Estimates for industrializing New England run even higher.

By the mid-nineteenth century, the use of children, especially in industry, became controversial. Governments imposed age restrictions and other measures that reduced children’s earning opportunities. The British “Factory Acts,” starting in 1833, imposed restrictions on the ages of children who could work, and how many hours they could work. But they started at a modest level; the 1833 Act restricted children aged 9–12 to forty-eight hour weeks. Prussia introduced similar measures in 1839, with other German states soon following (Wehler 1996, p. 257). Many governments tied restrictions on child labor to an education requirement. The English Factory Acts required that child workers also be in school. In some cases, the factory had to set up its own school to continue employing children (Nardinelli 1990, pp. 106–07). The Prussian 1839 Act established a minimum work age of nine years, and sixteen years for children who had not yet had at least three years of schooling. In Massachusetts as of 1837, manufacturers could not employ anyone under the age of 15 who had not attended school at least three months in the previous year (Carolyn Moehling 1999, p. 74).

There are two styles of explanation for the new child-labor laws. One is that a combination of social-welfare concerns, along with representatives of labor concerned about competition with adult males, overwhelmed industrialists’ opposition to child-labor restrictions. The other explanation is that these measures were enacted when industry no longer opposed them; either it had become easy to substitute capital and other sorts of labor for child labor, or the workforce had already changed in ways that the new laws were not a binding constraint when passed.

Child-labor restrictions potentially reduce the incentive to have a large family, but we have to bear in mind their limitations. Most such measures either did not apply to agricultural work, or did so in a more relaxed way. Wehler (1996) emphasizes that the German restrictions did not successfully limit the role of children in production at home, which remained important throughout the nineteenth century. And in every case, the restrictions’ impact would depend both on enforcement measures and parents’ desire to evade them. Finally, if child-labor restrictions were introduced when they were mostly irrelevant, then they could not be a strong causal force in the fertility transition.

4.4 Increases in the Opportunity Costs of Childbearing

Industrialization usually changed the role of women in the workforce, although economic historians do not agree on just how. Several studies show that women played important roles in factory work early in the
industrialization process, but became a less important part of that labor force as time went on. In a few industries such as textiles, women were certainly quite numerous into the twentieth century. Sara Horrell and Jane Humphries (1995, table 1) estimate wives’ labor-force participation rates of about 65 percent for England in the period 1787–1815, which corresponds to the height of the Industrial Revolution. This figure appears to fall in the late nineteenth century. Most other accounts report women as a proportion of the workforce in particular industries. In Britain in 1851, women constituted about thirty percent of the country’s measured labor force, and about forty percent of all employed women worked in textiles and clothing (Duncan Bythell 1993, p. 35). Claudia Goldin and Kenneth Sokoloff (1982, table 3) estimate that women comprised 20–30 percent of the workforce in New England textile factories in 1820, and about twice that in 1832. Gary Saxonhouse and Gavin Wright (1984, table 1) report that 57.6 percent of the workforce in cotton textiles in the American South in 1880 were female. Wehler (1996, p. 254) reports that women could be half the workforce in some German factories in the 1830s and 1840s; in 1875, according to the census, women constituted nearly half of the workforce in textile and clothing factories in Germany, and about twenty percent of industrial workers overall.

Married women certainly worked prior to the Industrial Revolution, but industrial work created new opportunities and trade-offs for women. It offered better-paying work that could not be combined with child-minding: a woman spinning yarn at home could also care for children, while a woman working in a textile factory could not. Some industries refused to hire married women at all, thus giving women an incentive to delay marriage. Industrialization thus raised the opportunity cost of children in two ways.

Several studies find that local employment opportunities for women lowered fertility. N. F. R. Crafts (1989) relies on the fertility and occupational information in the 1911 census of England and Wales. He finds a consistent, negative correlation between women’s local labor-force opportunities and marital fertility, with elasticities ranging from –0.13 to –0.34. Studies such as Brown and Guinnane (2002), which uses both textile mills and the structure of local agriculture to proxy for women’s earnings opportunities, find small, statistically significant effects with the expected signs. Schultz (1985) uses a different approach that links women’s earnings opportunities to the fertility transition per se. Using time series-cross section data on Swedish counties for the period 1860–1910, he shows that the ratio of women’s to men’s wages explains about a quarter of the decline in Swedish fertility. He treats women’s earnings as endogenous, and instruments for them using demand-side shocks to agricultural prices. Women’s earnings depress fertility at virtually all ages, so this effect seems to work through more than delayed marriage.

30 Historical censuses do a poor job of reporting women’s occupations, especially married women’s occupations. Horrell and Humphries (1995) base their study on family budgets.
31 Jürgen Kocka (1990, table 16). This figure refers only to establishments with five or more employees.

32 The “Lowell system” used by some textile mills in New England before the Civil War recruited young women to work in the mill and live in a special company boardinghouse under the supervision of a “housemother.” The point was to recruit farmers’ daughters who would otherwise be unwilling to undertake factory work (Saxonhouse and Wright 1984, pp. 4–5).
33 Marianne H. Wanamaker (2010) uses the introduction of textile mills in South Carolina in the period 1880–1900 to study the impact of changes in opportunity costs on family sizes. The introduction of a textile mill reduced fertility in the surrounding area by about 11 percent c. 1900. The effect reflects differential migration of low-fertility couples and so does not reflect a shift in the opportunity cost of childbearing in the sense of Becker’s model.
4.5 Changes in the Costs of and Returns to Child Quality

Another appealing idea is that fertility decline reflects increases in the net return to child quality. There are two different questions to ask. First, did the cost of child quality, in the form of education, decline? Second, did the return to child quality increase? A positive answer to either question implies a substitution away from numbers towards child quality. Unfortunately we can say more about the former than the latter.

Economic historians of education stress important distinctions in the types of economically useful education. One could acquire basic literacy and numeracy at home (if the parents were literate) or in primary school. More advanced education or training required secondary schools, formal apprenticeship, or on-the-job training. Tertiary education during the relevant period was restricted to a small elite, and while perhaps important for overall TFP growth, would not figure heavily in demographic decisions.

The growth of literacy and its primary cause, public elementary education, differs dramatically across the countries on which we focus. There are two broad classes of important decisions: First, to make primary schooling universally available, and second, to make it compulsory. Prussia led the way with the 1763 requirement that all children aged five to thirteen attend primary schools. The schools were not free, but there was tuition assistance for the poor. Like many grand educational reforms, this measure’s implementation was resisted by various interests, and in any case Prussia lacked sufficient teachers for all the children in the territory (James van Horn Melton 1988, pp. 174–77). In 1816, about 60 percent of students required to attend school actually did so, a figure rising to 82 percent by 1846. Some German states, such as Saxony, did better (Nipperdey 1994, p. 463). Several U.S. states introduced free public elementary education starting in the 1840s and, for most of the nineteenth century, the United States was an outlier in the proportion of young people in school. Free, compulsory education came later to Germany (1872), France (1882), and England and Wales (1893) (Kristine Bruland 2003, pp. 160–61). Easterlin (1981, appendix table 1) estimates that in 1850, there were 1,500 children in primary schools per 10,000 total population in the United States, compared to 1,600 in Germany, 930 in France, and 1,045 in the United Kingdom.

Literacy is both a broadly comparable measure of educational outcomes and the most economically useful output of primary schools. David F. Mitch (2004, tables 12.5A and 12.5B) reports that, in the time around 1860, adult male illiteracy rates were about 35 percent in France, 30 percent in England, 30 percent in the United States (whites only), and 5 percent in Prussia. Female illiteracy rates in these countries were higher: 45 percent for France, 37 percent for England, 10 percent for the United States (whites only), and 5 percent in Prussia. Female illiteracy rates in these countries were higher: 45 percent for France, 37 percent for England, 10 percent for the United States (whites only), and 5 percent for Prussia. The rapid development of schools in both Britain and France dramatically increased literacy rates by the end of the nineteenth century (François Furet and Jacques Ozouf 1977, p. 293).

The creation of schools certainly reduced the cost of elementary education. But the opportunity cost of the time spent in school remained, even when schools were free. Mitch (1992, p. 156) quotes Horace Mann’s toward subjects such as religion. Primary school students learned to read and write, and some basic arithmetic, but more advanced skills were acquired in other ways (Thomas Nipperdey 1994, p. 462).
comment in the British 1851 census of education: “It is not for sake of saving a penny per week, but for the sake of gaining a shilling or eighteenpence a week that a child is transferred from the school to the factory.” Child-labor laws might have been more important for encouraging schooling than the schools themselves.

What were the returns to education? Historical sources usually report only literacy status, not years of education, and occupation rather than income or wage. So we cannot estimate returns to education as is in the modern literature on the economics of education. Goldin and Lawrence F. Katz (2000) exploit the unique Iowa census of 1915 to provide one of the only historical estimates available. They find that the return to an additional year of high school or college then was, for males, on the order of 11–12 percent (Goldin and Katz 2000). Mitch (1992, pp. 230–35) estimates the present value of acquiring literacy in Victorian Britain for a representative child. The present value of the cost of acquiring literacy would be about £4. At a wage premium of 5 shillings per week for literacy, the present value of the higher wages for a 35-year work life would be over £200.

Efforts to examine the relationship between children’s education and their parents’ fertility also confront data problems. Sources usually do not include both parental fertility and children’s educational attainment, which is what we need for direct tests of the Q–Q trade-off. It would be tempting to interpret parental education as a proxy for children’s education. But this kind of effect can be interpreted in many different ways. The paper on Prussia discussed earlier (Becker, Cinnirella, and Woessmann 2009) takes the preferred approach, which is to estimate the impact of school enrollment rates on fertility. Presumably this approach could be replicated in other contexts, using changes in schooling and child-labor rules to measure the net returns to education.

4.6 Social Insurance and Old-Age Support

One particular return to child-rearing that receives considerable attention in the literature is children’s role in insuring parents against the consequences of accidents, ill-health, and old age. The most common argument is that children are a form of life-cycle savings; parents invest when they are young and healthy, and then expect their children to care for them in infirmity or old age. (The Sundstrom–David criticism of Easterlin’s model, noted above, is one version of this argument.) We should also consider insurance against accidents and ill-health. Two versions of this argument have been advanced to explain the fertility transition. One is that industrialization and the increased mobility it entails, especially rural-to-urban migration, made it harder for parents to hold children to the intergenerational bargain. This “child default” argument implies that the developing industrial economy made children a less desirable vehicle for savings. A second version of the argument points to the development of substitute means of providing for old age, especially social insurance and the welfare state.

Both versions of the argument suffer from the problem that we know that economic ties between parents and children varied dramatically across the societies in question before the fertility transition. In some European regions, peasant households would draft formal documents that turned a farmstead over to the heir, and carefully specified the heir’s obligations to his parents (as well as to siblings to who had not yet received an inheritance). This Altenteil guaranteed specific transfers to the retired couple, who often lived on the farm in a special house reserved

This would also be consistent with Caldwell’s changes in “net intergenerational wealth flows.” Carter, Ransom, and Sutch (2004) invoke the child default argument in discussing the U.S. fertility transition.
for that purpose. At the other extreme, rural laborers’ children in England would, from at least the early-modern period, leave home for good in their early to mid teens. The best evidence suggests these children had no further economic relationship with their parents (Alan MacFarlane 1986, pp. 83–84). Thus the extent to which parent–child ties changed during the nineteenth century varies a great deal. We should also note that the “child default” version of the argument resembles arguments about mortality (although in reverse): from the parents’ viewpoint, rising child default is like increased infant and child mortality. Parents’ might actually invest in more, lower-quality children to ensure that at least some children remained faithful.

This argument faces a different kind of challenge, which is that the social-insurance systems introduced at the end of the nineteenth and early twentieth century were usually replacing earlier schemes. Thus there is no clear “before.” Prior to the introduction of social insurance, every society in question here had some form of provision for the poor and those unable to provide for themselves because of illness or age. These “poor relief” systems were locally financed and organized, and were rarely as generous as the social-insurance schemes that eventually replaced them. Another important difference was their logic: poor relief systems were intended to relieve “need,” and so imposed asset tests. The numbers receiving relief at any one time were small, as one would expect. E. P. Hennock (2007, pp. 46–47) estimates that 6.6 percent of the English population and about 3.4 percent of the German population received poor relief in 1885. Social insurance, on the other hand, generally operated on insurance principles: covered individuals received specified payments triggered by specified events or conditions.

The literature divides social insurance into four categories: sickness insurance, (workplace) accident insurance, old-age and disability insurance, and unemployment insurance. The first broad social-insurance system dates to 1883, when the German government introduced compulsory sickness insurance (1883) and then accident insurance (1884). Disability and old-age insurance were added in 1889, while unemployment insurance came later. The system’s introduction was less discrete than these dates imply. Some German workers were already covered by schemes that were compulsory for their industry, and that became the model for the system for which Bismarck is always credited. The sickness and accident-insurance programs covered only workers in selected industries at first, although coverage broadened rapidly. Participation in the disability and old-age insurance program were limited by income. The United Kingdom’s Old Age Pensions Act (1908) illustrates a different approach. This noncontributory scheme was introduced for all persons who met a (mild) means test and had reached the age of 70 years. France and other Continental

37 Many workers in the late nineteenth century were also covered by voluntary, private schemes that provided assistance in case of sickness, accident, or infirmity. Some individual employers and labor groups also created insurance programs that covered parts of the workforce. In Germany, some workers were obliged to join a particular insurance organization long before the introduction of formal social insurance. Guinnane and Jochen Streb (2011) discusses Friendly Societies and the German organizations. John E. Murray (2007) discusses these organizations more broadly and provides references to the broader literature.

38 David Khoudour-Castéras (2008) reports that the health insurance law covered 21 percent of workers in 1885, rising to 44 percent in 1913. The accident insurance system at first covered 18 percent of workers, but by 1913 covered 94 percent (appendix tables 1 and 2).
countries tended to emulate the German approach to social insurance, introducing first sickness and accident insurance, usually in the 1890s or later. There were earlier old-age pension systems, but they were voluntary and did not receive state subsidies, and were not widely used. The United States was an outlier: before the Social Security Act of 1935, the United States had no social insurance per se. Local poor relief systems, war pensions, and mother’s pensions filled some of the role met by poor relief and social insurance in Europe. The “workman’s compensation” system introduced in the United States in the early twentieth century filled the role of German-style accident insurance (Price V. Fishback and Shawn Everett Kantor 2000).

The timing of social insurance’s rise in Europe hints that it is part of the story of the fertility transition. We cannot identify a single date at which social insurance first appeared, however. On the other hand, with care one might track the extension of such programs to different parts of a national population, and look for the impact on fertility of changes in social insurance over time. The broad patterns also do not make it likely that social insurance alone is central to the story. The two forerunners, France and the United States, were laggards in developing social insurance.

5. Conclusions

The historical fertility transition played a central role in the making of modern economies. This paper outlines the central empirical patterns in a selection of important, wealthy countries, and then provides an overview of the major economic explanations for the fertility transition. I caution at several points that apparently minor differences in demographic measures can make an important difference to the patterns economists seek to explain. More generally, much of the relevant literature in this area was produced by demographers and others who sometimes use definitions that to economists are unfamiliar and perhaps surprising. These differences reward care in consuming that literature.

There are several different economic explanations for the historical fertility transition. Two doubtless played a role, but are easy to exaggerate. During the late nineteenth century technological change introduced the first widespread use of “modern” contraceptives such as condoms. Yet the imperfect evidence available suggests these methods were still expensive, and that most couples continued to rely on “traditional” methods such as abstinence and withdrawal. Furthermore, simulation models show that traditional methods were sufficiently effective to account for the fertility decline we observe. Another often-stressed explanation for fertility decline is the decline of mortality, especially infant and child mortality. Mortality decline probably played a role in the historical fertility decline, but I have stressed two caveats. In some countries, fertility declined significantly before any real mortality decline. In addition, infant and child mortality are at least partially endogenous to fertility.

The other explanations discussed here all work off relative price changes, as implied by Becker’s demand-for-children model. Several significant changes in the relevant period plausibly altered the costs of and returns to children in ways that would reduce fertility. These include housing costs due to urbanization, changes in child-labor law, increases in the opportunity costs of childbearing because of better labor-force opportunities for women, the introduction of free or compulsory primary education, and the development of social-insurance systems.

Despite at least one hundred years of academic and official interest in the decline of fertility, this question is not one for which economists have a clear, empirically
well-founded explanation. (This is not to say there are not many partial answers based on particular times or places, or many theoretical efforts that capture some important part of the process.) There are several reasons for the current state of the literature. The most important is extensive simultaneity: the relevant period saw many significant economic changes, and some of these changes were tied to other changes such as the introduction of social insurance. Economists are well aware (and more recently, working hard on) the interrelations between fertility and the economy. But there remains the problem of understanding the role of any given change in a period when many other important forces changed at the same time. This is of course partly a standard identification problem, and partly an issue of being able to invoke ceteris paribus reasoning. Both require much more careful empirical work than we presently have.

Much more empirical work is both necessary and possible. The recent literature typically relies on national aggregates or, in some cases, subnational aggregate data. This approach sometimes reflects a reasonable trade-off between the costs and benefits of data collection. But for many historical economies there remain significant possibilities for using data much better-suited to testing the hypotheses described. The most important unexploited sources are individual-level data, especially individual-level data that include wealth and income, or reasonable proxies for wealth and income (such as occupation). The latter point is crucial: we know relatively little about the historical fertility transition because much of the earlier research relied on sources that do not contain useful economic information. Individual-level information is important for several different reasons. Sometimes data of this sort allows the researcher to follow life-cycle fertility, which permits use of a broader range of shocks to identify behavioral responses. Even when the sources are cross-sectional, individual-level data supports tests of multivariate hypotheses. (With aggregate data, on the other hand, there are usually ecological inference problems with any multivariate hypothesis.) These empirical approaches require significant investment of researcher time and other resources, but hold out the promise of really understanding, for example, the effect of women’s labor-force opportunities on the demand for children during the fertility transition. The recent upsurge in theoretical interest in the historical fertility transition will pay an even greater dividend if it motivates renewed interest in using the available data to understand the economics of the historical fertility transition.

REFERENCES


This article has been cited by: