Economics 719
Mira (2007)
December 13, 2007

I'd like to say a few words on Mira's 2007 *IER* paper. The substantive topic is on the fertility response to child mortality. From a demographic standpoint, in most societies the transition from a high–mortality/high–fertility regime to a low–mortality/low–fertility regime occurred with the decline in mortality preceding the decline in fertility. From a practical standpoint this pattern yields rapid population growth. The qualitative nature of the change is robust (mortality declines before fertility) but there is wide variation across countries in magnitudes (the size of the declines) and timing (i.e., the length of the lag).

Understanding the behavioral mechanisms involved is important for predicting future population trends. And for assessing the value of different kinds of mortality reduction. Deaton argues that declines in old–age mortality increase social welfare, while declines in infant mortality may not. The ambiguity hinges on whether parents have another birth to replace an infant that dies. If there is full replacement, then reducing infant mortality reduces the number of births but no increase in longetivity.

There have been a variety of approaches to estimate the replacement response. T. Paul Schultz and Ben–Porath introduced two strategies. They start with a premise that parents' have preferences over the **surviving** number of children. Strategy one is replacement – to replace a child that dies. Alternatively, parents may "hoard" children — have more than optimal in anticipation that some will die. Hoarding behavior is likely to be more important if mortality rates are high for school age or teenage children. If so, then it is possible that a child may die after the woman's reproductive ages and can not be replaced. Hoarding behavior is simply holding an inventory. There is no need to hold a positive inventory if there are no production lags and current demand can always be satisfied.

Generally, however, infant mortality rates are high and then mortality rates decline quickly and remain low through young adulthood. In the US and other developed countries, mortality rates increase especially for males in the late teen age years — when they can drink and drive.

Mira's paper is thus a contribution to this substantive literature. He extends the analysis published in Ken Wolpin's 1984 JPE paper. Wolpin's paper is one of the first to present estimates of a structural dynamic discrete choice model. Wolpin's paper offered a structural representation of the fertility decision making process. He uses this representation to estimate the prevalence of replacement behavior. Prior work used various (and sometimes ingenious) descriptive approaches to estimate replacement behavior. Wolpin finds little replacement in the Malaysian data, roughly half that reported in the descriptive literature (.15 versus .30).

Wolpin modeled the mortality regime as uncertain, where the probability of an infant death is $p(x_t)\exp(\epsilon_t)$, where $p(x_t)$ is the deterministic component (x_t can include time varying variables (calendar time) and individual characteristics, in principle) and ϵ_t is an iid stochastic component. Infant mortality rates may be time varying, but are exogenous to the individual.

Mira extends Wolpin by assuming that women have an unknown, individual specific, mortality rate. There is unobserved heterogeneity in mortality rates, and women can learn about their level through the survival status of their children. So Mira embeds the decision problem in a Bayesian learning framework. And it on learning and its modeling that I want to focus.

To see the impact, write a simple Bellman equation with explicit notation for the woman's information set at time t, I_t. Then Bellman's equation becomes:

$$
\begin{aligned}
V_t(x) &= \max_a \{u(x,a) + \beta E[V_{t+1}(x')|I_t]\} \\
&= \max_a \{u(x,a) + \beta \sum_{x'} V_{t+1}(x') f(x'|x,a,I_t)\}
\end{aligned}
$$

The point is that the expectation of future payoffs, is taken with respect to the woman's subjective beliefs. Because she is learning about her true mortality level, the transition probabilities will depend on her history. Otherwise identical women with the same number of children, but a different number of surviving children at time t will have different subjective beliefs about the law of motion.

And the modeling challenge here is that we need to keep I_t to low dimension if we will have a feasible model. Remember our interest is in **Markov** strategies. Because with Markov decision processes we only need to keep track of the current state, and not the path that took us to that state.

With the latent infant mortality risk, dynamic programming problem is one of imperfect state information. In the absence of unobserved heterogeneity, agents know completely their state, and apply their decision rule $\sigma(x, \epsilon)$. They are uncertain about x' but I assume they see the disturbance in period t before making the decision in t.

The true value of her mortality risk is unknown to her. This makes the model more complex but the solution is straight forward. **Ask:** What's the solution?

Yes, that's right. We need to augment the state space by sufficient statistics that summarize the information set. The expectations are taken with respect to the subjective distribution. In a model of Bayesian learning, we integrate against the posterior distribution, determined from by the observed signals and history that comprise the information set. Thus, we see that in models with imperfect state information we have have to add sufficient statistics necessary to characterize the posterior distribution to the state space.

As we know, in models of bayesian learning simplicity is gained by working with conjugate families. For example, since the mortality process is a bernoulli process, $f(d|p) = p^d(1-p)^{1-d}$, the prior density for p, $\pi(p)$, is the beta distribution. The posterior density, $\pi(p|d)$ is also a beta distribution but with parameters reflecting the observed outcome.

There is a huge computational simplicity if only infant mortality is permitted. That is, if a child survives the first period, he or she will survive through the entire decision period. Otherwise if child mortality rates are positive then the state space must include each child's age. Not so bad for one child, but it is easy to see how the state space grows with 5 kids, as the five can be any combination of ages.

It is important to realize that $f(x'|x, a, I)$ represents subjective beliefs. To close the model, we have to specify the learning rule. Bayesian learning is an obvious one, but not the only rule. Equally obvious and even simpler is the – no learning rule, myopia. $f(x'|x, a, I) = f(x'|x, a)$, subjective beliefs do not depend on the past history. Equally common is another no learning model, which is perfect foresight. There is no learning because everything is known about the path of mortality. The actual mortality experience is unrealized and thus unknown. So, if f is shifting over time, the shift is fully anticipated.

Using one's history is also only one form of learning. The social interactions literature also studies social learning. In this case, a woman's assessment of the mortality risk would depend on the outcomes for her kids, but also the mortality experience for others known to her, other family members or other women in the village.

Now things get interesting. How to model the learning from neighbors? One feature of Bayesian learning, is that people learn too fast; agents remember too much. If a woman sees the outcomes for herself and say n neighbors she's getting $n + 1$ signals every period. Subjective beliefs will converge quickly.

Yet that's for Bayesian learning. In agent base learning models, simple learning rules are postulated and their implications explored, with no presumption the rules are "optimal." So, in the village example, assign $p = p_B$ if 3 of the 11 infants die, and other assume $p = p_G$ (B for Bad and G for Good). The idea is that we will only revise our assessment if we observe an extreme outcome. By Bayes rule this can not be optimal because it's wasteful; we're not using the information for usual, less extreme outcomes.

What makes Malaysia an interesting application, is that the infant mortality rates declined substantially (by about 2/3) during the sample period. This raises the question, is what is the critical or important learning problem? Learning about one's own risk factor or forecasting the mortality decline? Mira assumes perfect foresight on the secular decline. The only uncertainty is how the individual's risk varies about the common, forecastable, trend. Mira states on page 820:

> Modeling the process by which women obtain their priors and learn about the trend as well as about individual–specific components is beyond the scope of this paper.

And footnote 20 continues

> Learning about the trend in the presence of individual heterogeneity is difficult because the effect of trend itself and the effect of the selectivity of fertility need to be disentangled in the aggregate data. But this cannot be done without knowledge of all individuals' decision rules and the distribution of the population across the state variables. So it seems that a model incorporating learning about mortality trends would involve a very complex fixed point problem in the space of decision rules.

I agree, but why is it complex? What is the selectivity of fertility? To learn about the current regime, the agent will have to use information from others. But the timing of fertility depends on each person's assessment of their risk. People with high survival chances are more likely to have kids at younger ages (i.e., with declining mortality those with high risk find it cost effective to postpone fertility). So need a subjective estimate of the population distribution of mortality risk to determine who has a kid when. And with rational expectations, realized probabilities equal expected probabilities which gives rise to the fixed point in strategy space.

However, this is for rational expectations and for solving the joint problem of disentangling the trend from individual heterogeneity. Alternatively, one can assume there is no individual heterogeneity, there is only secular trend. And solve the forecasting problem.

And with social learning, I would expect there should be regional or spatial differences in mortality rates. I can conjecture learning rules, and then work out how I would expect them to manifest themselves in what I observe. This provides some descriptive tests before estimating the hard DP problem.

I will close on a caveat. I know infant mortality rates decline from 15 to 5 (or lower) percent. I haven't been able to determine from the paper, how much that changes the strategy on fertility. The counterfactual I have in mind, take two identical women. Let one face a constant mortality risk of 15 percent and the other 5 percent, with prices and incomes as they are. How different are the fertility profiles? If essentially other factors are driving fertility, learning may be a very small part of the problem and not worth detailed analysis.