# Economic Development as Problem Solving: A model of Guns, Germs and Steel

Yaw Nyarko
New York University

May 2007

**Abstract**

I thank Boyan Jovanovic for many helpful comments. I also thank Silvana Melitsko for superb research assistance. I thank the audiences at many seminars and conference meetings for very helpful comments.

## 1 Introduction

There has been a vast recent literature which has emphasized the role of human capital and knowledge on the economic growth process. This paper presents a model where the knowledge occurs through solving "problems." These problems which are partially idiosyncratic to the country. We think of a problem as a new technique or good. Each country must master the problems associated with an activity in order to achieve productivity increases and growth. We will follow the tradition of learning by doing models where learning occurs after production on a good, or a particular type or grade of a good. Growth occurs by solving or learning about successively harder problems.

What is emphasized in this paper are two issues. First, as is well known, countries can improve their productivity and learn to produce new goods by observing other nations, especially as they solve their own production related problems. Solutions to one country's problem, however, only translate partially help solve one's own country's problems. The growth-enhancing effects of learning from others depends upon how closely related (or correlated) are the problems of one country with those of the others. Second, we introduce the concept of the "direction" of growth in our theoretical model. We consider a situation where countries could choose a number of different directions of growth. They choose the ones that are best for them, given their individual circumstances.

As regards first issue addressed by this paper, the learning from other countries, we are able to model a hypothesis which was described eloquently in the recent best-seller by Jared Diamond "Guns, Germs and Steel:" the East-West versus North-South Axes Hypothesis. Consider figure 1 below[1]:

Figure 1:

A grossly simplified version of this hypothesis says that Europe developed faster than Africa because Europe has a wide East-West axis which allowed for a lot of shared learning among similar countries, an advantage lacked by Africa because of its lying on a North-South axis.

On the East-West axis, the geography of countries is similar, so things learned or problems solved at one location are easily transferred to another location on the east-west axis. On a North South Axis however, learning at one site is not readily transferred to another site, because the geography quickly changes - in the African example, from dense forest at the equator, to savannah vegetation and then finally to desert in the North. Not only is the geography different, but perhaps there are natural barriers which will stop the movement of people - the desert for example may restrict movement of people and ideas from the Northern coast of Africa to the forest regions. This paper will provide a model where, under some parameter values, this Axes hypothesis will be true.

It must be empasized that the primary purpose of this paper is theoretical - providing a framework for modeling and analyzing knowledge transfer, learning and economic growth. Although we appeal to existing empirical assertions (e.g., the Jared Diamond hypothesis just mentioned), this paper does not take a stand on

[1] I thank Prof. Ramon Marimon for this figure.

these assertions. We merely observe that the model presented here can capture the empirical assertions.

Even if one accepts the Axes hypothesis, one need not obtain the conclusion that the implicit information sharing is always an unmitigated good thing. On the contrary, we will explore situations we will describe as a knowledge trap. In this situation, learning from others may distort the direction of learning and lead to the poorer countries remaining poor and for long run growth to be less than it could have been if learning from others was absent. Because the leading countries are solving their own problems and sharing the solutions with the followers, the followers skew their production and learning in the direction of the leaders. The followers drop some "unsolved" production activities in favor of those solved by the leaders. This results in the followers remaining followers forever.

We also point out that under some situations the leading countries could be better in the long run if the followers had not dropped some of those activities. Indeed, World optimality may require that the poor not be left too far behind on the technological ladders and, further, that the followers continue with production which is idiosyncratic to and better for the followers .

The paper proceeds as follows. In section 2, we begin by reviewing the model with only one country. This will be taken from Jovanovic and Nyarko (1996). Within-country learning takes place after production, so may be considered learning by doing. In section 3 we then introduce learning from others. In particular, countries observe the learning by doing signals of others. However, since no two countries are the same, the learning by doing transferred will be less, the less similar (or correlated) the countries. Further, if two countries are observing essentially similar signals (i.e., highly correlated) then the information content of those signals is small. After discussing the learning from neighbors model, we discuss our main hypotheses outlined earlier: the Jared Diamond hypothesis, the knowledge trap, and trade distortions.

It may be useful to relate this paper to the Jovanovic and Nyarko (1996) model which developed the learning by doing Bayesian model used here. That paper focussed on a one-country model, with little attention to sharing of information among countries. A principal insight of that paper was that high human capital countries may be at a disadvantage because they would have an incentive to stick to their high human capital grade of technology, while a low human capital country would have an incentive to upgrade to newer technologies, thereby eventually overtaking the human capital country. The knowledge trap we develop in this paper, on the other hand, the high human capital country remains a high human capital, and causes relatively lower growth in the other countries by distorting those countries growth paths in the direction of the high human capital country.

In this paper, we think of the units of study as nations. In principle, we could of course think of our units as firms. Our analysis would therefore be related to firms which have many to copy from as opposed to those who have only a few. Characterizations of learning by individual firms using a set up related to this paper

can be found in the very nice papers by Kleenow (1997) and Mitchell (2000).

### 1.1 Other Applications and Related Literature

This issue of a knowledge trap is related to the debate on foreign aid and economic development of poorer countries. When outsiders advise locals on how to organize their internal economies, there are two issues related to learning. First, by solving problems for locals, the outsiders may be preventing the locals from benefiting from learning by doing. Second, the outsiders may be proposing solutions which work in their own countries, and which they are the most familiar with. With this external learning on outsider's technologies, the locals may, ex post, find it optimal to develop in the direction proposed by the outsiders. Under some parameter values this will lead to lower long term growth in comparison to if the locals had been allowed to develop their own economies using local methods.

This issue of the knowledge trap is of course related to work on international trade, the infant industry argument and dynamic comparative advantage. Some representative papers in this literature include Arrow (1962) , Clemhout and Wan (1970), Bardhan (1971), Hoff (1997), Matsuyama (1992), Lucas (1988) and Young (1991). The concept of directions of economic growth used in this paper is related to the literature on trade strategies and learning in Rodrik (2005) and Hausmann and Rodrik (2003).

Related to this we ask about (i) learning taking place through trade; (ii) whether comparative advantage leads countries to solve the same or different problems and whether it leads to different countries copying each other.

We could work the same thing through consumption goods. By learning technologies which are based on outsiders tastes, the locals decide to produce for their own market consumption goods which are those of the outsiders.

Further, trade itself could lead to specialization in the direction of the leaders. The basic idea is that demand in the leading countries could skew the production and hence the learning in the follower countries.

## 2 The One-Country One-Good model

A technological line is made up of grades indexed by n$\in$ $[0,\infty)$. A country must choose a grade n to use, and must also choose a decision variable z representing how to operate that grade, resulting in a net output given by

$$q = \gamma^n \left[1 - (y_{nt} - z)^2\right], \quad \gamma > 1 \tag{1}$$

where

$$y_{nt} = \theta_n + w_{nt}. \tag{2}$$

Here $y_n$ is the target for the decision variable z, $\theta_n$ is the parameter relating to grade n; $w_{nt}$ is a zero mean which is normal with zero mean and variance $\sigma_w^2$. The agent does not know the value of $\theta_n$ but has a prior which $\mathcal{N}\left(E_t\theta_n, var_t\left(\theta_n\right)\right)$. If grade n is chosen, the risk neutral agent seeking to maximize expected net output would set

$$z = Ey_{nt} = E\theta_{nt} \tag{3}$$

resulting in expected output

$$E_t q = \gamma^n \left[1 - var_t\left(\theta_n\right) - \sigma_w^2\right]. \tag{4}$$

We will define the human capital s to be the reciprocal of the variance, s$\equiv 1/x$, sometimes also referred to as the precision. After working on grade n, the country will experience learning by doing which will reduce its variance to $h_1\left(x\right)$. Specifically, the country will observe y in (2) which, from Bayesian updating, results in a posterior human capital $\mathcal{H}_1\left(s\right)$ and variance $h_1\left(x\right)$ given by

$$\mathcal{H}_1\left(s\right) = s + \frac{1}{\sigma_w^2} \quad \text{and} \quad h_1\left(x\right) = \frac{\sigma_w^2 x}{\sigma_w^2 + x}. \tag{5}$$

(For emphasis, note that for s=1/x, we have $\mathcal{H}_1\left(s\right) = 1/h_1\left(x\right)$.) We suppose that the parameters of any two grades n and n+k for k>0 are related by the following:

$$\theta_{n+k} = \alpha^{k/2}\theta_n + e_k \tag{6}$$

where

$$e_k \tilde{} N\left(0, \rho_k\sigma_\varepsilon^2\right) \quad \text{and } \rho_k = \begin{cases} \left(1-\alpha^k\right)/\left(1-\alpha\right) & \text{for } \alpha \neq 1 \\ k & \text{for } \alpha\text{=1} \end{cases} \tag{7}$$

The above is the generalization of the AR-1 process to a diffusion process.

It will be useful to define the following two functions:

$$\mathcal{H}_2\left(s,k\right) = \frac{s}{\alpha^k + s\rho_k\sigma_\varepsilon^2} \quad \text{and } h_2\left(x,k\right) = \alpha^k x + \rho_k\sigma_\varepsilon^2 \text{ and} \tag{8}$$

$$\mathcal{H}\left(s,k\right) = \mathcal{H}_1\left(\mathcal{H}_2\left(s,k\right)\right) \text{ and } h\left(x,k\right) = h_1\left(h_2\left(x,k\right)\right) \tag{9}$$

If s is the human capital of grade n, then the function $\mathcal{H}_2\left(s,k\right)$ (resp. $\mathcal{H}\left(s,k\right)$) gives the human capital on grade $\theta_{n+k}$ before (resp. after) learning by doing on grade n+k. Define $s_k^{**}$ to be the fixed point of $\mathcal{H}\left(.,k\right)$ (one can show that it exists and is unique, and that iterates of $\mathcal{H}\left(.,k\right)$ from any s converge to $s_k^{**}$). In particular,

if there is a jump of size k in each period followed by learning on that grade, the posterior human capital will converge to $s_k^{**}$. We will define

$$\hat{x} = \frac{\sigma_\varepsilon^2}{1-\alpha} \text{ for } \alpha \neq 1 \text{ and } \hat{s} = \frac{1}{\hat{x}} \quad . \tag{10}$$

Note that $\hat{s}$ is the value of s such that $\mathcal{H}(s,k) = s$, and will play a key role in our analysis below.

Define G$(x,k)$ to be the expected net output when x is the variance on the current or status quo grade and a jump of size k is chosen:

$$G(x,k) = \gamma^k \left[1 - \sigma_w^2 - \alpha^k x - \rho_k \sigma_\varepsilon^2\right]. \tag{11}$$

We write a production function F in terms of s= $\frac{1}{x}$ and the grade k:

$$\mathrm{F}(s,k) = G\left(\frac{1}{s}\right). \tag{12}$$

The production function F is increasing and strictly concave in s.[2]

We suppose that once a grade has been passed for a higher grade, it is never recalled.

Define $s^*$ to be the value of s such that F$(s,0) = F(s,1)$. We now impose some conditions on our parameters, which we refer to as Assumption B, ensuring that here is long run positive growth from all initial conditions. Jovanovic and Nyarko (1996) identified two cases. Case A was characterized by the conditions $\alpha\gamma < 1$; $F(\infty,0) > F(\infty,1)$ (or, equivalently, $G(0,0) > G(0,1)$) and $s^* < s_1^{**}$. In case A, initial human capital conditions affect long-run growth catastrophically - low human capital s (high x) countries have long run growth while high human capital s (low x) countries always use the same grade and hence never grow. In case B, which we use here but refer to as Assumption B, there is long run positive growth from all initial conditions.)

**Assumption B**: (i) $\alpha\gamma > 1$; (ii) for some k>0, 0< F$(\infty,0) < F(\infty,k)$ (or, equivalently, G$(0,0) < G(0,k)$) and $x^* < x_1^{**}$; and (iii) if $\alpha < 1$ then G$(\hat{x},0)$ <0.

Parts (i) and (ii) of the above assumption were used in Jovanovic and Nyarko (1996) with k=1, and deliver the conclusion that there are spurts of growth (upgrading) followed by spurts of only learning by doing on a given grade. Condition (iii) is a technical condition required to rule out the possibility that the optimal jump size is $+\infty$ from all human capital levels[3].

[2] One feature of F is that for very low s (high x) the production function is negative. This feature will neither drive nor be important to the conclusions of this paper.

[3] This follows from the observation that when $\alpha < 1$, $\lim_{k\to\infty} \frac{1}{\gamma^k} G(x,k) = 1 - \sigma_w^2 - \hat{x}$.

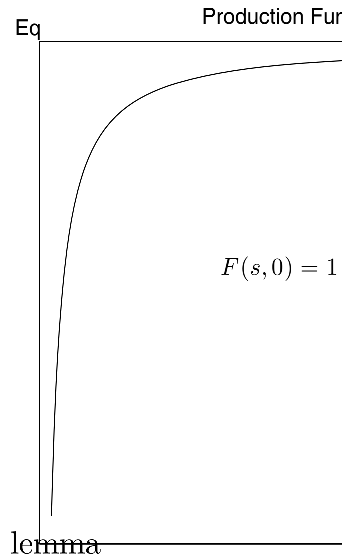

Figure 2: The production function

Define $\kappa^* (s)$ to be the optimal jump size from human capital s. The lemma below indicates that from each s there is a unique optimal jump size, and further that the jump size is increasing in the human capital s. The lemma is stated after the presentation of some figures illustrating what the output and optimal upgrading under our parameters.

As can be seen from the figure, the higher is the precision, the higher is the grade that should be chosen. The production function figure shows that at low values of the precision, lower upgrading is optimal. In comparing Hong Kong and Singapore, Alwyn Young (1992) suggested that Singapore was upgrading at too high levels, suggesting that lower values of upgrading would have resulted in higher returns, as indicated by the production function above.

**Proposition 1** *Suppose Assumption (B) holds. Then (i) $\kappa^* (s)$ is uniquely defined and finite from each s and (ii) $\kappa^*$ is increasing in s; in particular, for $s > s'$, $\kappa^* (s) \geq \kappa^* (s')$, with strict inequality if $\kappa^* (s') > 0$. More specifically,*

$$\kappa^* (s) = Max \left\{ \bar{\kappa}^* (s), 0 \right\} \quad \textit{where} \tag{13}$$

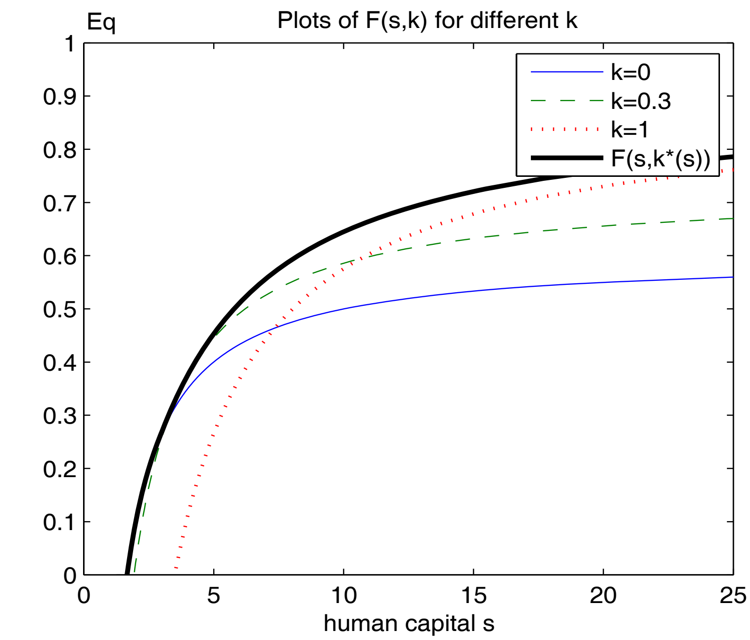


Figure 3: Plots of F(s,k) for different k

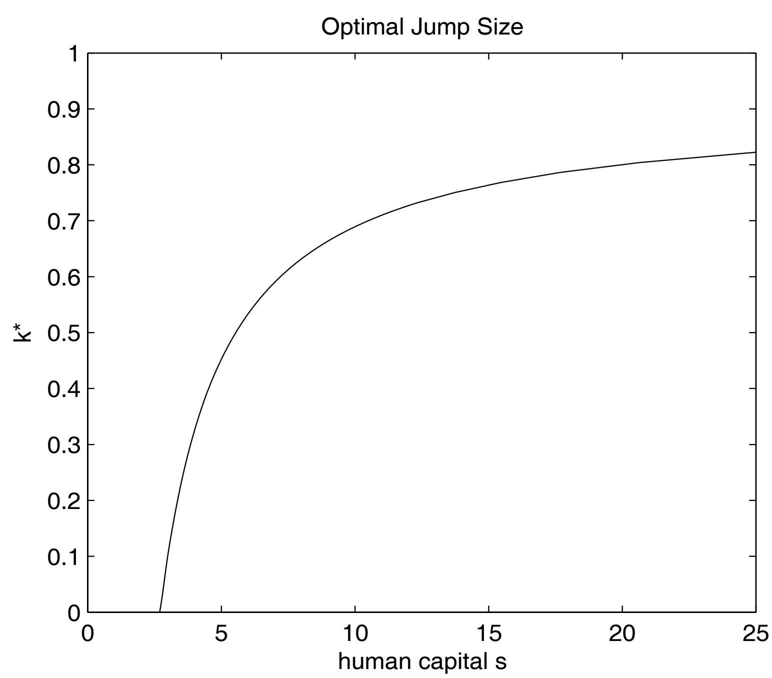


Figure 4: The optimal jump (upgrading)

*where*

$$\bar{\kappa}^*(s) = \begin{cases} \frac{1}{\ln\alpha}\ln\left(\frac{-(\ln\gamma)G(\hat{x},0)}{\ln(\alpha\gamma)\left(\hat{x}-\frac{1}{s}\right)}\right) & \textit{for} \quad \alpha \neq 1 \\ \frac{(\ln\gamma)F(s,0)-\sigma_\varepsilon^2}{\sigma_\varepsilon^2\ln\gamma} & \textit{for } \alpha = 1 \end{cases}. \tag{14}$$

*Proof of Proposition* (1) See Appendix.

# 3 Learning from Neighbors

How does learning from neighbors affect dynamics? Here we will study a model where there is symmetric information among all agents - everybody has the same aggregate information. However, since nations apply this information to different activities the information is differentially effective in the different locations. The symmetry of information makes this different from the bulk of models with learning studied in the relevant literatures, where the differences are in the information received, and not, like here, in the uses of the common information.

Under the conditions of this section, knowledge spillovers may assist in growth. Two countries may each be growing slowly when they are on their own, but with information sharing may decide to upgrade faster so long as there is some correlation in the parameters of their grades. This will be the mechanism we will use to capture the diffusion of ideas between different countries.

Our analysis will be as follows. On their own, countries will have a certain path of growth - meaning technological upgrading behavior. With information sharing there will be superior information, and also different growth paths. We shall look at the possible growth paths as a function of our key parameters $(\alpha, \gamma, \sigma_w^2, \sigma_\varepsilon^2)$ and also as a function of the correlation between the signals of different countries, which will be captured by a parameter $\rho$.

## 3.1 The Correlation Structure Among Nations

For each nation, the optimization problem within a period is exactly the same as that described in the earlier section. The only difference between this section and the earlier one is in the learning by doing, which now occurs not only from the nation's own efforts, but also those of other countries. The information from others will be useful because the parameters of production - the $\theta's$ - of each nation are correlated[4].

In particular, let $\Psi$ denote both the set of countries as well as the number of countries. Each country has a technological line with different grades indexed by

[4] We assume that the learning by doing noise terms, $\{w_t^c\}$ are independent across nations. We will later on take up the situation where the correlation is in the error terms of the learning by doing component.

n$\in$ $[0,\infty)$. The parameter of country $c\in\Psi$ is denoted by $\theta_n^c$ . At the beginning of each date country $c$ must choose a grade n$^c$ and a decision variable z$^c$ on the grade, resulting in an output q$^c$, as in equations (1) - (4).

Let

$$\bar{\boldsymbol{\theta}}_n = \{\theta_n^c\}_{c\in\Psi} \tag{15}$$

be the vector of grade n parameters of each country. We suppose that the elements of $\bar{\boldsymbol{\theta}}_n$ are correlated and, in particular, have a joint normal distribution with a variance-covariance matrix X which has the symmetric form

$$X = xI(\rho_\theta) \;\; where \;\; I(\rho_\theta) = \begin{bmatrix} 1 & \rho_\theta & ... & \rho_\theta \\ \rho_\theta & 1 & ... & \rho_\theta \\ ... & ... & ... & ... \\ \rho_\theta & \rho_\theta & ... & 1 \end{bmatrix}, \; x>0 \;\; \text{and}^5 \; \rho_\theta \in [0,1]. \tag{16}$$

In particular, we suppose that there is a common variance $x>0$ of each $\theta_n^c$ and a common correlation coefficient $\rho_\theta$ between any $\theta_n^c$ and $\theta_n^{c'}$ for c$\neq c'$.

Fix a date and let n$_c$ be the grade operated by country c at that date. After each country has operated on their chosen grade, each country will observe the vector of signals from all countries, their own and others, represented by the following vector:

$$Y_n = \left\{y_{n_c}^c\right\}_{c\in\Psi} \qquad \text{where} \qquad y_{nt}^c = \theta_n^c + w_{nt}^c. \tag{17}$$

In order to get some explicit results, we will impose some symmetry conditions. Suppose that at the very first period each country c has a some mean and a common variance x on the parameter of their first grade, $\theta_1^c$. Suppose further that the grade 1 parameters of any two parameters $\theta_1^c$ and $\theta_1^{c'}$ share a common correlation coefficient $\rho_\theta$. Further, impose on the countries common values of all the other relevant parameters of the model $(\gamma, \sigma_\varepsilon^2, \; \sigma_w^2)$.

Then from the optimization problem of the earlier section it should be clear that each country will choose the same optimal upgrading behavior in the first period. Since they receive the same information, they will all update their beliefs and have identical values for the variance on the parameters of their own grades. They will then repeat the optimization problem and again choose the same grades to operate on in the next, and hence each subsequent period. Each country will be on the same grade at each date, under the above mentioned conditions.

After observing the signals of all nations, each country c will update their beliefs about their own parameter, $\theta_n^c$. We let

$$X' = h_1(x, \Psi, \rho_\theta) \tag{18}$$

denote the updated or posterior variance covariance matrix after observation of the $\Psi$ signals Y$_n$ whose correlation structure is defined in via $(x,\rho_\theta)$ in (16). A

comparison with the one country model indicates that $\mathrm{h}_1(x)$ of the earlier section is the one country version of $h_1(x, \Psi, \rho_\theta)$ of this section. The lemma below indicates that $X'$ will have a form similar to (16). Further, part (2) of the lemma has the key result we will use in later - namely that the posterior variance is decreasing in the correlation coefficient $\rho_\theta$ between nations.

**Lemma 2** *(Posterior Distribution)*

***1.*** *The posterior variance covariance matrix is given by*

$$h_1(x, \Psi, \rho_\theta) = x' I(\rho'_\theta) \tag{19}$$

*where* $x' > 0$ *and* $\rho'_\theta \in [0,1]$, *and where, in particular,*

$$x' = \frac{\sigma_w^2 + x\left[1 + (\Psi - 2)\rho_\theta - \rho_\theta^2(\Psi - 1)\right]}{\frac{\sigma_w^2}{x} + \frac{x}{\sigma_w^2}\left[1 + (\Psi - 2)\rho_\theta - \rho_\theta^2(\Psi - 1)\right] + (\Psi - 2)\rho_\theta + 2} \tag{20}$$

*and*

$$\rho'_\theta = \frac{\rho_\theta \sigma_w^2}{\sigma_w^2 + x\left[1 + (\Psi - 2)\rho_\theta - \rho_\theta^2(\Psi - 1)\right]}\ . \tag{21}$$

***2.*** *Furthermore,*

$$\frac{\partial \rho'_\theta}{\partial \Psi} < 0, \quad \frac{\partial \rho'_\theta}{\partial \rho} > 0, \quad \textit{and}\ \frac{\partial x'}{\partial \Psi} < 0 \tag{22}$$

*and*

$$\frac{\partial x'}{\partial \rho} \quad \textit{is} \quad \begin{cases} < 0 & for \quad \rho > 0 \\ = 0 & for \quad \rho = 0 \end{cases}. \tag{23}$$

**Proof:** See Appendix 1.

If $s = \frac{1}{x}$ is the prior HK, then the posterior HK will be $s' = \frac{1}{x'}$ where $x'$ is given by the lemma above.

**Example 3**

In the above, we look at a model where there are two identical countries. We consider at three situations: autarky ($\rho_\theta = 0$), moderate correlation ($\rho_\theta = 0.5$) and complete correlation ($\rho_\theta = 1$),and plot the posterior human capital $s'$ as a function of the prior human capital s.

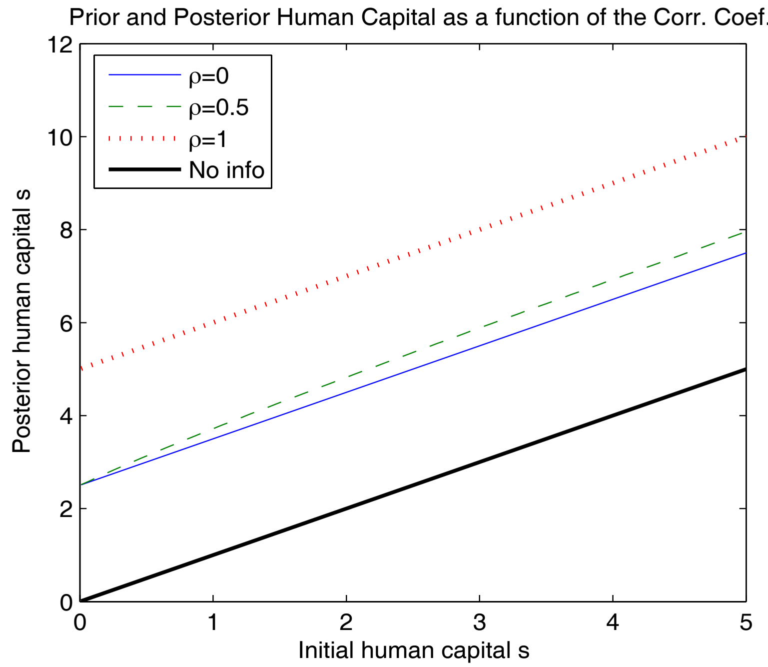


Figure 5: Prior and posterior human capital

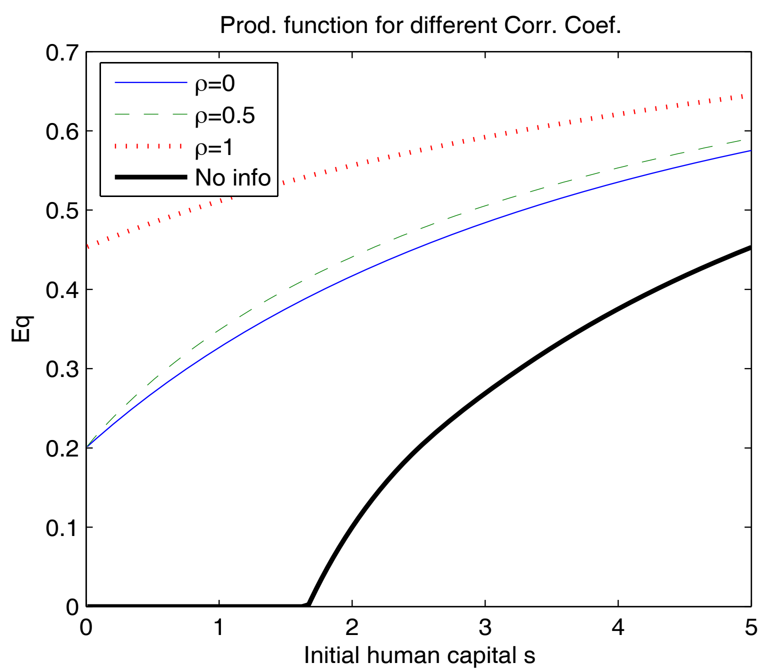


Figure 6: Production as a function of $\rho_\theta$

**Remark 4** *As $\rho_\theta$ goes to 1, we should approximate the model with many signals on the same $\theta$. This is easily seen to be the case since from* (20) *above we obtain*

$$\lim_{\rho_\theta \to 1} s' = s + \frac{\Psi}{\sigma_w^2} \tag{24}$$

*with the right hand of the above being the posterior variance with $\Psi$ independent signals on a common parameter $\theta_1$.*

**Remark 5** *Eq.* (22) *above also shows that as $\Psi$ increases, the posterior variance decreases - this captures the fact that signals are better for information processing. Further, as $\Psi$ increases the posterior correlation coefficient between any two different $\theta'$s decreases.*

**Remark 6** *It is interesting to note the limit of $s'$ as $\Psi \to \infty$ :*

$$\lim_{\Psi \to \infty} s' = \frac{1}{\sigma_w^2} + \frac{s}{1-\rho}. \tag{25}$$

*As $\rho_\theta \to 1$, $s'$ becomes the posterior human capital (precision) on a given $\theta$ after getting information on parameters which are almost perfectly correlated. As $\Psi \to \infty$, we would therefore expect the posterior human capital to go to $\infty$. The formula above shows this.*

## 3.2 The Optimal Dynamics, Growth Paths and the Jared Diamond Hypothesis

In the lemma of the sub-section above, we stated conclusions in terms of the final human capital: in particular, we noted that the higher is the absolute value of the correlation coefficient between two nations the higher is the human capital after information sharing. We re-state this here as a proposition.

**Proposition 7** *The (common) posterior human capital is increasing in $\rho_\theta$.*

We interpret the correlation coefficient to be the measure of informational distance between nations. A high correlation coefficient may model the situation where nations are on an axis with similar geography and therefore where information sharing is easy. The above proposition could therefore be a statement, in terms of human capital, of the Jared Diamond hypothesis - those on axes more conducive to knowledge transfer (in his theory, the East-West axis), will have higher human capital after knowledge transfer.

One may ask whether there is a direct translation between the correlation coefficient and the growth rates over time. After all, it is not human capital per se that we are interested in, but instead in the long run growth. In particular, we ask the question: can we extend the analysis of the earlier section by providing a link between higher human capital and higher growth rates? When there is such a link, we obtain the "Jared-Diamond" hypothesis in terms of growth rates.

Shortly, we will provide conditions under which the higher human capital does indeed result in higher growth rates. Before we do this, however, it is important to note that there need not necessarily be a positively monotonic mapping from higher human capital to higher growth rates. There are two kinds of reasons for this. The first reason was stressed in Jovanovic and Nyarko (1996), who noted that it is possible that a high human capital situation may lead one to specialize in that activity for which there is higher human capital, resulting in a reluctance to upgrade, and a lower growth rate than a low human capital country with no stake in the current or status quo technology. The conditions for this are in what we described as Case A earlier. Our Assumption B puts rules out this case, and instead implies that higher human capital (lower variance) leads to higher growth rates in the period.

There is a second, more subtle, reason for the possibility of the lack of a positive monotonic relationship where higher human capital maps into higher growth rates. This situation could arise under Assumption B, which is the case we are working with. As indicated in the earlier lemma, in case B, higher human capital results in a higher jump in technologies. A country with higher human capital would therefore have a higher growth *in the period* of the jump. The mapping from jump size into human capital on the new grade is monotonically decreasing, however. Hence in the following period the country that did the higher jump in the first period would have a lower human capital on their chosen grade, and hence a lower jump size in the subsequent period. In particular, it is possible that the country which grew the fastest in the initial period may grow slower in the next period and in many subsequent ones. Figuring out long run growth rates will therefore not be straightforward in this case.

On the other hand, it is always easy to generate some conditions under which the higher correlation coefficients translate unambiguously into higher growth. To see this let us begin with an extreme case. Suppose that $\Psi$, the number of countries, is large, and that $\rho_\theta^A = 1$ and $\rho_\theta^B = 0$. Since $\rho_\theta^A = 1$,the observation of the signals of others provides for each country in Group A $\Psi$ independent signals on their own $\theta$. Since by assumption $\Psi$ is very large, the countries in Group A will have practically complete information on their own value of $\theta$ after observing the learning by doing signals of others; in particular the variance on their status quo grade at each date t is $\mathrm{x}_t^B = 0$. The countries in Group B, however, with $\rho_\theta^B = 0$ will not have any use for the signals of others since these are uncorrelated with their own $\theta$. They will therefore have a positive variance $\mathrm{x}_t^B > 0$ on their own $\theta$ at the end of each period. Since the jump size is monotonically decreasing in the variance we see in this case that countries in group A will always have a jump size $\mathrm{k}^{*A} = k^*(0)$ which will always

exceed that of countries in group B.

By the obvious continuity of the model it should be clear that we can always find a non-trivial set of parameters where this conclusion is true for all periods t or at least for a large number of periods.

More formally we may state the following:

**Proposition** Fix any time horizon T<$\infty$, however large. Let $\rho_\theta^A \in [0,1]$ and $\rho_\theta^B \in [0,1]$ be any two correlation parameters. Consider two groups of countries, Group A and B, who differ only by their correlation coefficients, $\rho_\theta^A$ and $\rho_\theta^B$ respectively. Let $k_t^A$ and $k_t^B$ denote their optimal jump sizes. Then so long as $\Psi$ is sufficiently large and $\left(\rho_\theta^A - \rho_\theta^B\right)$ is sufficiently close to 1,

$$k_t^A \geq k_t^B \;\text{ for all t.} \tag{26}$$

In particular, in each and every period the Group I countries will have a higher technological jump size than those in Group B and, in particular, the Jared Diamond East West Hypothesis is satisfied and the lower (resp. higher) the correlation between the parameters associated with the technologies of the countries, the lower (resp. higher) is the growth rate.

**Remark 8** *Of course, in the above proposition we have imposed somewhat stringent conditions on $\rho_\theta^A$ and $\rho_\theta^B$ to obtain the fairly stringent conclusion that the jump sizes in* each *period can be ordered. We could of course weaken the conditions to obtain weaker conclusions which would imply for example that the average jump sizes over time can be ordered.*

## 3.3 Some Comments and Remarks:

What is a country or nation? We have modeled above a country as the entity where there is a "unit" of learning by doing in each period. If a country is big, we could in principle think of there being multiple sources of independent learning by doing. This distinction will of course become important in any empirical implementations of the model. A unit could be a nation if we think of a legal system with property rights where inventions acrue to one person or firm, with other independent inventions on the same activity unlikely.

There is another issue related to that just raised. In the earlier sections we have assumed that the $\theta's$ differ across countries, but that countries get *independent* signals on their individual $\theta's$. We concluded in this case that the more correlated are the $\theta's$,the better is the learning and hence growth. We could assume, instead,

that the $\theta's$ are the same but the signals (the w's) themselves are correlated. If the signals of countries are, for e

It would then be expected that the effect of correlation will be the opposite of what we have for the earlier case: higher correlation between the noise terms of the signals will imply lower learning and hence lower growth. In particular, in a model with both correlated $\theta's$ and correlated error terms of the signals w, the effect of increased correlation works in opposite directions.

Going back to the Jared Diamond East-West hypothesis, this adds a layer of subtlety to the basic argument. On the one hand, those on an East-West axis may have much more correlated $\theta's$ (which is good for growth), they may also be getting similar signals on the similar $\theta's$, which will be bad for growth. It is easy to extend the analysis above to handle this case, and indeed in Appendix 2 we sketch some of the algebra and the footnote at the end of the appendix formalizes the possibilities mentioned in this discussion.

# 4 Empirical Evidence For Model

The goal of this paper is primarily theoretical and not empirical. Our principal purpose is to indicate how our model can shed some light on some empirical assertions made in the literature, and in particular the assertions by Jared Diamond (1997) alluded to in the introduction. There are a number of empirical papers in the economics literature, however, which we believe provide partial justification for the basic modeling here. We mention some of these works - they are the ones we have laid our hands on, or have been suggested. This is not meant to be an exhaustive survey of the related empirical literature. We provide a series of papers whose empirical conclusions seem to go in the direction of the spirit of the model presented here.

*1.* Alwyn Young (1992) studies the basic trade-off between learning more and thereby getting more productive on an existing activity as opposed to upgrading to a potentially better activity that you know less about. This was applied in a discussion of the relative growth experiences of Singapore and Hong Kong. That paper illustrates that fact that the upgrading behavior may be one chosen by a government (the Singapore case), which may be thought of as "forced upgrading" or in a more laissez faire manner (the Hong Kong case).

*2.* Easterly and Levine (1998):"Troubles with Neighbors: Africa's Problems, Africa's Opportunity," argue that contagion of policies from one country to its neighbor may be a very important explanatory variable in explaining growth outcomes, and in particular Africa's growth outcome. In growth regressions they show that the standard variables (human and physical capital, good policies, etc), on their own "account for half of the growth rate differential between slow growing Africa and fast growing East Asia." They find that the growth rate in one country is highly correlated with the growth rate of another country. They find evidence that supports the

conclusion that neighboring countries imitate each others' policies. This copying, they hypothesize, may either be copying of good policies following successful government interventions or policies which maximize rent-seeking behavior by government not necessarily maximizing growth rates. They hypothesize as one of the channels of the imitation and contagion of policies situations where "local conditions are similar among neighbors... ." Easterly and Levine justify their hypotheses via their very interesting empirical work. As should be clear, these hypotheses are the very building blocks of the model of this paper. Here, we model information sharing as the mechanism through which copying and imitation take place. In our model, imitation is much easier among countries close together (neighbors) than those far away.

*3.* One can view our model as providing a possible model to explain some of the empirical data mentioned in the literature of Club Convergence In that literature (see Baumol, W. J. (1986) and Durlauf, S. and P. A. Johnson (1995)), it is claimed that countries are in different clubs, with similar behavior within clubs and different behavior outside of clubs. In our model, the different axes play the same role as clubs in the earlier mentioned literature.

*4.* There is also a recent literature on putting climate and geography (and in particular percentage of country in tropical region) into growth regressions (e.g., Bloom and Sachs (1998), Sachs (2001, 2003)). In these models the axes and directly and literally entered into regressions.

# 5 Directions of Growth

Suppose that there are two nations A and B. There are also two technological lines, I and II, and each nation has access to them. We suppose that each nation can choose only one of these technological lines to operate on at a time. We will refer to these technological lines as different directions of growth.

We use the term informational autarky to represent the situation where there is no information sharing among the nations. We will provide an example of a situation where under informational autarky, each nation chooses the technology which is the best for them, whereas under informational sharing, on the other hand, nation B will copy A and do the upgrading using the technology which is inferior for B, simply because A has used it. Country B, by copying A ends up always a laggard nation, and in the long run produces an output which is less than what it would produce in the short run.

Suppose that there are two technologies, I and II, in existence. We will suppose that the parameters of the two technologies are completely independent. Technology I dominates II for nation A, while the opposite is true for nation B (see 7). We suppose that when a country uses the technology which is superior (resp. inferior)

Figure 7: Directions of Growth and Knowledge Trap

for it, the growth parameter for that technology is $\lambda^H$ (resp. $\lambda^L$), where

$$\lambda^H > \lambda^L. \tag{27}$$

In particular, the superior technology in one nation is exactly the same as the inferior technology in the other. The productivities or payoffs of any given technology differs in the two countries because of a host of unmodeled factors like geography, climate, culture etc.

Suppose that each nation is at the steady state human capital level $s^{**} = \frac{1}{x^{**}}$ for their superior technology, and are at the same level for their inferior technolgy. Then under informational autarky, and assuming the conditions of the one country model above, each nation will be growing at the rate[6] $\lambda^H$.

Now suppose that one day the informational autarky between the nations ends, and there is information sharing between the two nations. In particular, suppose that country A produces at some period, and then the learning by doing signal on A's superior technology I is immediately relayed to country B. Suppose that the information from A to B leads B to revise the human capital precision up to a level $s'$. Suppose that $s'$ is sufficiently large and, in particular, $s' > \underline{s}$, where $\underline{s}$ is the value of s such that the output by Country B at $s^{**}$using the superior technolgy I is the same as the output by B at $\underline{s}$ using the inferior technology (see 7).

Then, upon receipt of the information from Country A on technology I, it becomes optimal for country B to use the inferior technology with the better human capital precision $s'$ than the superior technology with the worse human capital level $s^{**}$.

The long run growth rate of Country B then will be $\lambda^L$ with information sharing from Country A, rather than the higher level $\lambda^H$ which would occur under autarky.

For emphasis, we remark that example therefore has the following two properties:

1. Under informational autarky each country uses the technology which is best suited for it.

2. Under Informational sharing, the country that moves first uses the technology which is best suited to it, while the other country uses the technology which is not suited to it.

## 5.1 Comments

1. One may ask whether the knowledge trap requires a "leader" country and a "follower" country. This is clearly the interpretation implied by a modeling

[6]We normalize units so that $\lambda$ is defined such that in the steady state of the model each country chooses a jump size of 1 in each period. (To see explicitly how this is done, see Jovanovic and Nyarko (1996)).

- however note that we have made the two nations in many ways symmetric. Indeed, under the precise symmetric parameters used here, the country in the trap has initially a higher output level than the one being copied. This is seen in the graph on the time series of nations above.

2. Indivisibilities: The modeling here assumes that a nation can only operate on one technological line at a time. One would think that in reality, it may be possible to have two different technological lines in existence at the same time. We are thinking here, in the model, of situations where the new technological line completely eradicates the old: Western scientific medicine completely wipes out traditional medicine; new legal systems completely eradicate existing ones; etc.

3. Discount Factor: Clearly this model and the example holds for the low (or zero) discount factors. Clearly, if the discount factor of the nation goes to 1, then the future is all that is important, and the nation will obviously seek to maximize long run growth rates. What we show here is that when there is sufficient impatience, knowledge traps may be possible.

# 6 Efficient World Learning and Research and Development

The possibility of a knowledge trap illustrated above, suggests an interesting question: Is it possible that the knowledge trap could result in informational inefficiencies for world knowledge. Suppose that every now and then there is a remarkable innovation. Call this a bonus grade. This will be only useful if you have the skills or human capital to use it. If the world has not learned the technologies which the leader has decided are inferior, the leader may not be able to use the bonus grades when they become available. There are many who believe, for example, that herbal medicines hold many secrets useful for the world. Unfortunately, there are many tribal societies which will perish, with their knowledge, before the rest of the world will be able to get to them to learn their secrets.

# 7 Conclusion

We have provided a model which shows how learning by doing from differentially related countries can affect growth outcomes. We have also indicated how the optimal dynamics can lead to growth traps when learning by one country from another can

distort the direction of growth of the copying country. Some empirical justification for the modeling is provided.

# 8 References


1. Baumol, W. J. (1986): "Productivity Growth, Convergence and Welfare: What the Long Run Data Show?", American Economic Review, 76, 1072 – 85.

2. Baumol and Wolff (1988): "Productivity Growth, Convergence and Welfare: reply," American Economic Review 78, 1155-1159.

3. Bloom, D. and Sachs, J. D. (1998). 'Geography, demography and economic growth in Africa', Brookings Papers on Economic Activity, vol. 2, pp. 207–73.

4. Diamond, J. (1997): *Guns, Germs and Steel: The Fate of Human Societies*, W.W. Norton and Co., New York.

5. Durlauf, S. and P. A. Johnson (1995):"Multiple Regimes and Cross-Country Growth Behavior," Journal of Applied Econometrics, 10, 365-384.

6. Easterly, W. and R. Levine (1998): "Troubles with the Neighbours: Africa's Problem, Africa's Opportunity", Journal of African Economies, 7, 120-142.

7. Evenson, R.E. and L.E. Westphal (1995): "Technological change and technology strategy", in J. Behrman and T.N. Srinivasan, (eds.): *Handbook of Development Economics IIIA*. Amsterdam: Elsevier Science.

8. Hoff, K. (1997): "Bayesian Learning in an Infant Industry Model", Journal of International Economics, 43, 409-436.

9. Jovanovic, B., and Y. Nyarko (1996): "Learning By Doing and the Choice of Technology", Econometrica, 64, 1299-1310.

10. Jovanovic, B., and Y. Nyarko (1995): "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves", in *Brookings Papers on Economic Activity: Microeconomics,* 247-305, Brookings Institute, Washington, D.C.

11. Kleenow, P. (1998): "Learning Curves and the Cyclical Behavior of Manufacturing Industries", Review of Economic Dynamics, 1, 531-550.

12. Lucas, R. E. Jr. (1988): "On the Mechanics of Economic Development", Journal of Monetary Economics, 22, 3–42.

13. Matsuyama, K. (1992): "Agricultural productivity, comparative advantage, and economic growth", Journal of Economic Theory, 58, 317–34.

14. Mitchell, M. (2000): "The Scale of Production in Technological Revolutions", Research Department Staff Report 269, Federal Reserve Bank of Minneapolis.

15. Sachs, J. D. (2001). 'Tropical underdevelopment', NBER Working Paper W8119.

16. Sachs, J. D. (2003). 'Institutions don't rule: direct effects of geography on per capita income', NBER Working Paper W9490.

17. Young, A. (1992): "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore", in O. J. Blanchard and S. Fischer (eds.): *NBER Macroeconomics Annual, 1992*. Cambridge and London: MIT Press, 1992, 13-54.

# 9 Appendix 1: The Proofs

*Proof of Proposition* (1)

Simple algebra shows that,

$$G(x,k) = \begin{cases} \gamma^k \left[G(\hat{x},0) + \alpha^k (\hat{x} - x)\right] & \text{for } \alpha \neq 1 \\ \gamma^k \left[G(x,0) - k\sigma_\varepsilon^2\right] & \text{for } \alpha = 1 \end{cases} \tag{28}$$

so,

$$\frac{1}{\gamma^k}\frac{\partial G(x,k)}{\partial k} = \begin{cases} (\ln\gamma)\, G(\hat{x},0) + \alpha^k \ln(\alpha\gamma)(\hat{x} - x) & \text{for } \alpha \neq 1 \\ (\ln\gamma)\, G(x,0) - [1 + k\ln\gamma]\, \sigma_\varepsilon^2 & \text{for } \alpha = 1 \end{cases}. \tag{29}$$

When x$>\hat{x}$ and $\alpha < 1$, assumption B and (29) imply that $\frac{\partial G(x,k)}{\partial k} < 0$, so k$^*(x) =$ 0. The proposition therefore holds trivially for such x, and we therefore exclude this case from our discussion.

Recalling that when $\alpha > 1$, $\hat{x} < 0$ we see that from (29), that $\frac{1}{\gamma^k}\frac{\partial G(x,k)}{\partial k}$ is strictly decreasing in k for all k and is negative for all large k (and remember we are excluding the case x$>\hat{x}$ and $\alpha < 1$ which we have already dealt with).

This means that either $\frac{1}{\gamma^k}\frac{\partial G(x,k)}{\partial k}$ is everywhere negative, in which case k$^*(x)$=0; or else $\frac{1}{\gamma^k}\frac{\partial G(x,k)}{\partial k}$ is initially positive in which case $G(x,k)$ is initially increasing, reaches a maximum at k$^*(x) < \infty$, is thereafter decreasing. In either case, note that k$^*(x)$ is uniquely and well- defined.

Now we prove the monotonicity of k$^*(x)$ in x. Let $x < x'$ as in this lemma, and let k=k$^*(x)$ and k$'$=k$^*(x')$. If $k'$=0 then the lemma is trivial, so suppose that $k' > 0$ and further suppose, per absurdem, that k$\leq k'$.

Since $k'>0$, $\frac{\partial G(x',k')}{\partial k}=0$, so from simple algebra

$$\frac{1}{\gamma^{k}}\frac{\partial G(x,k)}{\partial k} = \frac{1}{\gamma^{k}}\frac{\partial G(x,k)}{\partial k}-\frac{1}{\gamma^{k'}}\frac{\partial G(x',k')}{\partial k} \tag{30}$$

$$= \begin{cases} \ln(\alpha\gamma)\left\{\alpha^{k}(\hat{x}-x)-\alpha^{k'}(\hat{x}-x')\right\} \text{ for } \alpha\neq 1 \\ (\ln\gamma)\left[G(x,0)-G(x',0)\right]-\sigma_{\varepsilon}^{2}(k-k')\ln\gamma \text{ for } \alpha=1 \end{cases} \tag{31}$$

We now indicate that the Right Hand Side (RHS) of the above is strictly positive. This would imply from the above that $\frac{\partial G(x,k)}{\partial k}>0$, which would contradict the optimality of x from k. When $\alpha>1$, -$(\hat{x}-x')>-(\hat{x}-x)>0$ so $0<\alpha^{k'}(\hat{x}-x')<\alpha^{k}(\hat{x}-x)$, and hence RHS $>0$. When $\alpha<1$, $0<(\hat{x}-x')<(\hat{x}-x)$ so $0<\alpha^{k'}(\hat{x}-x')<\alpha^{k}(\hat{x}-x)$ ,so RHS $>0$. When $\alpha=1$, $G(x,0)-G(x',0)>0$ and $k-k'\leq 0$ so again RHS $>0$.

**Proof of Lemma 1 (Posterior Distributions):**

For ease of exposition in the proof, use $\rho$ to represent $\rho_{\theta}$. Recall that $\mathbf{X}=xI(\rho)$. The Bayesian updating formulas can easily be verified to be given by

$$\begin{aligned} & (\mathbf{X}'')^{-1} \\ = \; & \mathbf{X}^{-1}+\frac{1}{\sigma_{w}^{2}}I(0) \\ = \; & \frac{1}{x}\left(\frac{1+(\Psi-2)\rho}{1+(\Psi-2)\rho-\rho^{2}(\Psi-1)}\right)I\left(-\frac{\rho}{1+(\Psi-2)\rho}\right)+\frac{1}{\sigma_{w}^{2}}I(0) \\ = \; & \hat{x}I(\hat{\rho}) \end{aligned}$$

where

$$\hat{x}=\frac{1}{x}\left(\frac{1+(\Psi-2)\rho}{1+(\Psi-2)\rho-\rho^{2}(\Psi-1)}\right)+\frac{1}{\sigma_{w}^{2}}$$

$$=\frac{\sigma_{w}^{2}\left[1+(\Psi-2)\rho\right]+x\left[1+(\Psi-2)\rho-\rho^{2}(\Psi-1)\right]}{x\sigma_{w}^{2}\left[1+(\Psi-2)\rho-\rho^{2}(\Psi-1)\right]}. \tag{32}$$

and

$$\begin{aligned} \hat{\rho} &= \frac{1}{\hat{x}}\left\{\frac{1}{x}\left(\frac{1+(\Psi-2)\rho}{1+(\Psi-2)\rho-\rho^{2}(\Psi-1)}\right)\left(-\frac{\rho}{1+(\Psi-2)\rho}\right)\right\} \qquad (33) \\ &= \frac{1}{\hat{x}}\left\{\frac{1}{x}\left(\frac{-\rho}{1+(\Psi-2)\rho-\rho^{2}(\Psi-1)}\right)\right\} \\ &= \frac{\sigma_{w}^{2}\left[1+(\Psi-2)\rho-\rho^{2}(\Psi-1)\right]}{\sigma_{w}^{2}\left[1+(\Psi-2)\rho\right]+x\left[1+(\Psi-2)\rho-\rho^{2}(\Psi-1)\right]}\left(\frac{-\rho}{1+(\Psi-2)\rho-\rho^{2}(\Psi-1)}\right) \\ &= \frac{-\rho\sigma_{w}^{2}}{\sigma_{w}^{2}\left[1+(\Psi-2)\rho\right]+x\left[1+(\Psi-2)\rho-\rho^{2}(\Psi-1)\right]}. \end{aligned}$$

Hence[7],

$$\begin{aligned}\mathbf{X}'' &= \frac{1}{\hat{x}}\left\{I\left(\hat{\rho}\right)\right\}^{-1} \\ &= \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) I\left(-\frac{\hat{\rho}}{1+(\Psi-2)\hat{\rho}}\right) \\ &= x' I\left(\rho'\right),\end{aligned}$$

where

$$x' = \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) \tag{34}$$

and

$$\begin{aligned}\rho' &= \frac{1}{x'}\left\{\frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right)\right\}\left(-\frac{\hat{\rho}}{1+(\Psi-2)\hat{\rho}}\right) \\ &= -\frac{\hat{\rho}}{1+(\Psi-2)\hat{\rho}} \\ &= \frac{\rho\sigma_w^2}{\sigma_w^2\left[1+(\Psi-2)\rho\right]+x\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right]-(\Psi-2)\rho\sigma_w^2} \\ &= \frac{\rho\sigma_w^2}{D},\end{aligned}$$

where

$$D \equiv \sigma_w^2 + x\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right].$$

This is of course the formula for $\rho'_\theta$ in $(21)$. We can therefore immediately obtain the derivatives of $\rho'_\theta$ with respect to $\Psi$ and $\rho$:

$$\frac{\partial\rho'}{\partial\Psi} = \frac{-\sigma_w^2}{D^2}\left[x\rho^2\left(1-\rho\right)\right] < 0.$$

---

[7] The computations use the following easily verified fact that if $I\left(\rho\right) = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{bmatrix}$. Then the inverse of $I\left(\rho\right)$ is as below:

$$\left(I\left(\rho\right)\right)^{-1} = QI(\rho').$$

where

$$\rho' = -\frac{\rho}{1+(N-2)\rho}$$

and

$$Q = \frac{1+(N-2)\rho}{1+(N-2)\rho-\rho^2(N-1)}.$$

and

$$\begin{aligned}\frac{\partial \rho'}{\partial \rho} &= \frac{\sigma_w^2}{D^2}\left[\left\{\sigma_w^2 + x\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right]\right\} - \rho x\left\{(\Psi-2)-2\rho(\Psi-1)\right\}\right] \\ &= \frac{\sigma_w^2}{D^2}\left\{\sigma_w^2 + x\left[1+\rho^2(\Psi-1)\right]\right\} > 0.\end{aligned}$$

which are respectively the first and second inequalities in (22).

Next we obtain expressions for $x'$ and its derivatives. Define

$$\Gamma \equiv [1+(\Psi-2)\rho] + \frac{x}{\sigma_w^2}\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right]$$

and

$$M \equiv \frac{x}{\sigma_w^2}\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right].$$

Then

$$\Gamma = 1+(\Psi-2)\rho + M.$$

Note from the expression for $\hat{\rho}$ in (33) that

$$\hat{\rho} = \frac{-\rho}{\Gamma}.$$

Hence from (32),

$$\begin{aligned}\hat{x} &= \frac{\sigma_w^2[1+(\Psi-2)\rho] + x[1+(\Psi-2)\rho-\rho^2(\Psi-1)]}{\sigma_w^2 x[1+(\Psi-2)\rho-\rho^2(\Psi-1)]} \\ &= \frac{\Gamma}{x[1+(\Psi-2)\rho-\rho^2(\Psi-1)]} \\ &= \frac{\Gamma}{M\sigma_w^2}.\end{aligned}$$

So from (34),

$$
\begin{aligned}
x' &= \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) \\
&= \frac{M\sigma_w^2}{\Gamma}\left(\frac{1+(\Psi-2)\left(\frac{-\rho}{\Gamma}\right)}{1+(\Psi-2)\left(\frac{-\rho}{\Gamma}\right)-\left(\frac{-\rho}{\Gamma}\right)^2(\Psi-1)}\right) \\
&= M\sigma_w^2\left(\frac{\Gamma+(\Psi-2)(-\rho)}{\Gamma^2+(\Psi-2)(-\rho)\Gamma-(\Psi-1)(-\rho)^2}\right) \\
&= M\sigma_w^2\left(\frac{1+M}{\Gamma^2+(\Psi-2)(-\rho)\Gamma-(\Psi-1)(-\rho)^2}\right) \\
&= \frac{M\sigma_w^2(1+M)}{\Gamma[\Gamma-(\Psi-2)(\rho)]-(\Psi-1)(-\rho)^2} \\
&= \frac{M\sigma_w^2(1+M)}{(1+(\Psi-2)\rho+M)[1+M]-(\Psi-1)(-\rho)^2} \\
&= \frac{M\sigma_w^2(1+M)}{1+(\Psi-2)\rho+2M+(\Psi-2)\rho M+M^2-(\Psi-1)(-\rho)^2} \\
&= \frac{M\sigma_w^2(1+M)}{2M+(\Psi-2)\rho M+M^2+\frac{M\sigma_w^2}{x}} \\
&= \frac{\sigma_w^2(1+M)}{2+(\Psi-2)\rho+M+\frac{\sigma_w^2}{x}}
\end{aligned}
$$

The above is the expression in (20). We may now take derivatives to obtain the following (where (*denum*) below represents the obvious denominator in the derivative, whose exact value is immaterial):

$$
\begin{aligned}
\frac{\partial x'}{\partial \Psi} &= \frac{\left\{2+(\Psi-2)\rho+M+\sigma_w^2/x\right\}\sigma_w^2\left\{x/\sigma_w^2\left(\rho-\rho^2\right)\right\}-\left\{1+M\right\}\sigma_w^2\left\{x/\sigma_w^2\left(\rho-\rho^2\right)+\rho\right\}}{(denum)^2} \\
&= \frac{\sigma_w^2\left\{x/\sigma_w^2\left(\rho-\rho^2\right)\right\}\left\{1+(\Psi-2)\rho+\sigma_w^2/x\right\}-\left\{1+M\right\}\left\{\sigma_w^2\rho\right\}}{(denum)^2} \\
&= \sigma_w^2\rho\frac{\left\{x/\sigma_w^2\left(1-\rho\right)\right\}\left\{\sigma_w^2/x+(\Psi-2)\rho+1\right\}-\left\{1+x/\sigma_w^2\left[1+(\Psi-2)\rho-\rho^2(\Psi-1)\right]\right\}}{(denum)^2} \\
&= \sigma_w^2\rho\frac{x/\sigma_w^2\{(1-\rho)\left[\sigma_w^2/x+(\Psi-2)\rho+1\right]-1-(\Psi-2)\rho+\rho^2(\Psi-1)\}-1}{(denum)^2} \\
&= \sigma_w^2\rho\frac{x/\sigma_w^2\{\sigma_w^2/x-\rho\sigma_w^2/x-(\Psi-2)\rho^2-\rho+\rho^2(\Psi-1)\}-1}{(denum)^2} \\
&= \sigma_w^2\rho^2\frac{-1+x/\sigma_w^2\{-(\Psi-2)\rho-1+\rho(\Psi-1)\}}{(denum)^2} \\
&= \sigma_w^2\rho^2\left(\frac{-1+x/\sigma_w^2\left\{\rho-1\right\}}{(denum)^2}\right),
\end{aligned}
$$

which is negative for $\rho > 0$ and 0 for $\rho = 0$, as in the lemma. Furthermore, if we let Z denote the obvious denominator term below,

$$
\begin{aligned}
\frac{\partial x'}{\partial \rho} &= \frac{\sigma_w^2\frac{\partial M}{\partial \rho}\{\frac{\sigma_w^2}{x}+M+(\Psi-2)\rho+2\}-\{1+M\}\sigma_w^2\{\frac{\partial M}{\partial \rho}+\Psi-2\}}{Z^2} \\
&= \frac{\sigma_w^2\frac{\partial M}{\partial \rho}\{\frac{\sigma_w^2}{x}+(\Psi-2)\rho+2\}-(1+M)\sigma_w^2(\Psi-2)-\frac{\partial M}{\partial \rho}\sigma_w^2}{Z^2} \\
&= \frac{\sigma_w^2\frac{\partial M}{\partial \rho}\{\frac{\sigma_w^2}{x}+(\Psi-2)\rho+1\}-(1+M)\sigma_w^2(\Psi-2)}{Z^2} \\
&= \sigma_w^2\frac{\frac{x}{\sigma_w^2}\{\Psi-2-2\rho(\Psi-1)\}\{\frac{\sigma_w^2}{x}+(\Psi-2)\rho+1\}-(1+M)(\Psi-2)}{Z^2} \\
&= \rho\sigma_w^2\frac{-2(\Psi-1)-2\frac{x}{\sigma_w^2}(\Psi-1)-\frac{x}{\sigma_w^2}\rho(\Psi-2)(\Psi-1)}{Z^2} \\
&= \sigma_w^2\rho(\Psi-1)\frac{-2-2\frac{x}{\sigma_w^2}-\frac{x}{\sigma_w^2}\rho(\Psi-2)}{Z^2}
\end{aligned}
$$

The expression is negative for $\rho > 0$ and 0 for $\rho = 0$. This completes the proof of the lemma. ■

# 10 Appendix 2: The Correlation Structure Among Nations via the learning by doing noise terms, w

Now we study the case where the noise terms for each nation are correlated. We follow an analogous process to that with correlation in the $\theta's$. In particular, fix a date t and let

$$\bar{\mathbf{w}}_n = \{w_n^c\}_{c \in \Psi} \tag{35}$$

be the vector of grade n signal errors of each country. We suppose that the elements of $\bar{\mathbf{w}}_n$ are correlated and have a joint normal distribution with a zero mean vector and variance-covariance matrix which has the symmetric form $\sigma^2_{\bar{\mathbf{w}}_n} = \sigma^2_w I(\rho_w)$ :

$$\bar{\mathbf{w}}_n \tilde{} \mathbf{N}\left(\mathbf{0}, \sigma_w^2 I(\rho_w)\right), \tag{36}$$

The signal error terms, $\bar{\mathbf{w}}_n$ at any date t are assumed to be independent of the vector at any other date $t'$ and at any other grade $n'$.

**Lemma 9** *(Posterior Distribution)*

1. *The posterior variance covariance is given by*

$$h_1(x, \Psi, \rho_\theta, \rho_w) = x' I(\rho'_\theta) \tag{37}$$

*where* $x' > 0$ *and* $\rho'_\theta \in [0,1]$, *and where, in particular,*

$$x' = \left( \frac{M(\rho_w)\sigma_w^2 + M(\rho_\theta)x}{2 + (\Psi-2)(\rho_w+\rho_\theta) - 2(\Psi-1)\rho_w\rho_\theta + M(\rho_w)\frac{\sigma_w^2}{x} + M(\rho_\theta)\frac{x}{\sigma_w^2}} \right) \tag{38}$$

*and*

$$\rho'_\theta = \Lambda\rho_\theta + (1-\Lambda)\rho_w \ . \tag{39}$$

*where we define*

$$M(\rho) \equiv 1 + (\Psi-2)\rho - (\Psi-1)\rho^2.$$

*and*

$$\Lambda = \frac{M(\rho_w)}{M(\rho_w)\sigma_w^2 + M(\rho_\theta)x}$$

**PROOF:**
Recall the definition of M($\rho$) above and let

$$Q(\rho) \equiv \frac{1 + (\Psi-2)\rho}{M(\rho)}.$$

Then one can verify that

$$
\begin{aligned}
& (\mathbf{X}'')^{-1} \\
= \;& \mathbf{X}^{-1} + \frac{1}{\sigma_w^2} I\left(\rho_w\right)^{-1} \\
= \;& \frac{1}{x} I\left(\rho_\theta\right)^{-1} + \frac{1}{\sigma_w^2} I\left(\rho_w\right)^{-1} \\
= \;& \frac{1}{x} Q\left(\rho_\theta\right) I\left(\frac{-\rho_\theta}{1+(\Psi-2)\rho_\theta}\right) + \frac{1}{\sigma_w^2} Q\left(\rho_w\right) I\left(\frac{-\rho_w}{1+(\Psi-2)\rho_w}\right).
\end{aligned}
$$

The elements in diagonal and off diagonal of $(\mathbf{X}'')^{-1}$ are

$$
\begin{aligned}
Diag &= \frac{1}{x} Q\left(\rho_\theta\right) + \frac{1}{\sigma_w^2} Q\left(\rho_w\right) \\
Offdiag &= \frac{1}{x} Q\left(\rho_\theta\right)\left(\frac{-\rho_\theta}{1+(\Psi-2)\rho_\theta}\right) + \frac{1}{\sigma_w^2} Q\left(\rho_w\right)\left(\frac{-\rho_w}{1+(\Psi-2)\rho_w}\right).
\end{aligned}
$$

We want

$$
(\mathbf{X}'')^{-1} = \hat{x} I\left(\hat{\rho}\right),
$$

which, from the diagonal elements, implies that

$$
\begin{aligned}
\hat{x} &= \frac{1}{x} Q\left(\rho_\theta\right) + \frac{1}{\sigma_w^2} Q\left(\rho_w\right) \\
\hat{x} &= \frac{1+(\Psi-2)\rho_\theta}{xM(\rho_\theta)} + \frac{1+(\Psi-2)\rho_w}{\sigma_w^2 M(\rho_w)} \\
\hat{x} &= \frac{(1+(\Psi-2)\rho_\theta)\sigma_w^2 M(\rho_w) + (1+(\Psi-2)\rho_w) x M(\rho_\theta)}{xM(\rho_\theta)\sigma_w^2 M(\rho_w)};
\end{aligned}
$$

and from the off-diagonal elements implies that ,

$$
\hat{x}\hat{\rho} = \left(\frac{-\rho_\theta}{xM(\rho_\theta)}\right) + \left(\frac{-\rho_w}{\sigma_w^2 M(\rho_w)}\right) = \left(\frac{-\rho_\theta \sigma_w^2 M(\rho_w) - \rho_w x M(\rho_\theta)}{xM(\rho_\theta)\sigma_w^2 M(\rho_w)}\right)
$$

so

$$
\hat{\rho} = \left(\frac{-\rho_\theta \sigma_w^2 M(\rho_w) - \rho_w x M(\rho_\theta)}{(1+(\Psi-2)\rho_\theta)\sigma_w^2 M(\rho_w) + (1+(\Psi-2)\rho_w) x M(\rho_\theta)}\right)..
$$

Hence,

$$
\begin{aligned}
\mathbf{X}'' &= \frac{1}{\hat{x}}\left\{I\left(\hat{\rho}\right)\right\}^{-1} \\
&= \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) I\left(-\frac{\hat{\rho}}{1+(\Psi-2)\hat{\rho}}\right) \\
&= x' I\left(\rho'\right),
\end{aligned}
$$

where

$$x' = \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) \tag{40}$$

and

$$\rho'_\theta = -\frac{\hat{\rho}}{1+(\Psi-2)\hat{\rho}} \tag{41}$$

After some manipulation one obtains the following formula for $\rho'_\theta$

$$\rho'_\theta = \frac{\rho_\theta M(\rho_w)\sigma_w^2 + \rho_w M(\rho_\theta)x}{M(\rho_w)\sigma_w^2 + M(\rho_\theta)x}$$

Since

$$
\begin{aligned}
Q\left(0\right) &= 1 \\
M(0) &= 1
\end{aligned}
$$

It is easy to see that this is consistent with all previous formulas:

$$
\begin{aligned}
\rho'_\theta(\rho_w &= 0) = \frac{\rho_\theta \sigma_w^2}{\sigma_w^2 + M(\rho_\theta)x} \\
&= \frac{\rho_\theta \sigma_w^2}{\sigma_w^2 + x(1+(\Psi-2)\rho_\theta - (\Psi-1)\rho_\theta^2)} \\
\rho'_\theta(\rho_\theta &= 0) = \frac{\rho_w x}{x + M(\rho_w)\sigma_w^2} \\
&= \frac{\rho_w x}{x + \sigma_w^2(1+(\Psi-2)\rho_w - (\Psi-1)\rho_w^2)}
\end{aligned}
$$

To find a formula for $x'$ define

$$
\begin{aligned}
\Upsilon &\equiv \left[1+(\Psi-2)\rho_\theta\right] M(\rho_w)\sigma_w^2 + \left[1+(\Psi-2)\rho_w\right] M(\rho_\theta)x \\
&= M(\rho_w)\sigma_w^2 + M(\rho_\theta)x + (\Psi-2)[M(\rho_w)\sigma_w^2\rho_\theta + M(\rho_\theta)x\rho_w]
\end{aligned}
$$

And

$$\tilde{M}(\rho_\theta) \equiv xM(\rho_\theta)\ ,\ \tilde{M}(\rho_w) \equiv \sigma_w^2 M(\rho_w)$$

So

$$\hat{x} = \frac{\Upsilon}{\tilde{M}(\rho_\theta)\tilde{M}(\rho_w)}$$

And

$$\hat{\rho} = \left(\frac{-\rho_\theta \tilde{M}(\rho_w) - \rho_w \tilde{M}(\rho_\theta)}{\Upsilon}\right)$$

Hence

$$\begin{aligned}
x' &= \frac{1}{\hat{x}}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) \\
&= \frac{\tilde{M}(\rho_\theta)\tilde{M}(\rho_w)}{\Upsilon}\left(\frac{1+(\Psi-2)\hat{\rho}}{1+(\Psi-2)\hat{\rho}-\hat{\rho}^2(\Psi-1)}\right) \\
&= C\left(\frac{\Upsilon+(\Psi-2)\hat{\rho}\Upsilon}{\Upsilon^2+(\Psi-2)\hat{\rho}\Upsilon^2-\hat{\rho}^2\Upsilon^2(\Psi-1)}\right) \text{where } C \equiv \tilde{M}(\rho_\theta)\tilde{M}(\rho_w) \\
&= C\left(\frac{\Upsilon+(\Psi-2)[-\rho_\theta\tilde{M}(\rho_w)-\rho_w\tilde{M}(\rho_\theta)]}{\Upsilon^2+(\Psi-2)[-\rho_\theta\tilde{M}(\rho_w)-\rho_w\tilde{M}(\rho_\theta)]\Upsilon-(\Psi-1)[-\rho_\theta\tilde{M}(\rho_w)-\rho_w\tilde{M}(\rho_\theta)]^2}\right) \\
&= C\left(\frac{\tilde{M}(\rho_w)+\tilde{M}(\rho_\theta)}{\Upsilon\{\Upsilon-(\Psi-2)[\rho_\theta\tilde{M}(\rho_w)+\rho_w\tilde{M}(\rho_\theta)]\}-(\Psi-1)[\rho_\theta\tilde{M}(\rho_w)+\rho_w\tilde{M}(\rho_\theta)]^2}\right) \\
&= C\left(\frac{\tilde{M}(\rho_w)+\tilde{M}(\rho_\theta)}{\Upsilon[\tilde{M}(\rho_w)+\tilde{M}(\rho_\theta)]-(\Psi-1)[\rho_\theta\tilde{M}(\rho_w)+\rho\tilde{M}(\rho_\theta)]^2}\right) \\
&= C\left(\frac{\tilde{M}(\rho_w)+\tilde{M}(\rho_\theta)}{\tilde{M}(\rho_w)^2M(\rho_\theta)+\tilde{M}(\rho_\theta)^2M(\rho_w)+\tilde{M}(\rho_w)\tilde{M}(\rho_\theta)(2+(\Psi-2)(\rho_\theta+\rho_w)-2(\Psi-1)\rho_\theta\rho_w}\right) \\
&\text{After some manipulation} \\
&= \left(\frac{M(\rho_w)\sigma_w^2+M(\rho_\theta)x}{2+(\Psi-2)(\rho_w+\rho_\theta)-2(\Psi-1)\rho_w\rho_\theta+M(\rho_w)\frac{\sigma_w^2}{x}+M(\rho_\theta)\frac{x}{\sigma_w^2}}\right)
\end{aligned}$$

**Remark 10** *The effect of changes in $\rho_w$ on the posterior variance is not a monotonic one. There are two general effects at work. First, suppose that $\rho_\theta = 0$. This is the case where the $\theta'$s of different countries are independent of each other. In this case, when $\rho_w = 0$, then one country's signal is irrelevant for the other. Now increase $\rho_w$ so that it is now positive. Then by observing the other country's signal, one gets some information the other person's w realization which is related to the own w. Hence the increase in $\rho_w$ from zero to a positive number provides information and in particular decreases the posterior variance $x'$. In particular, $x'$ is decreasing in $\rho_w$ in this case. On the other hand, now suppose that $\rho_\theta = 1$,so that we may think of there being two signals on the same $\theta$. In that case one would expect, and it is indeed the case, that having independent signals ($\rho_w = 0$) is best, and in particular the posterior precision increases as $\rho_w$ increases. For values of $\rho_\theta$ in between 0 and 1, the two effects are at work, and the effect of $\rho_w$ on the posterior variance is non-monotonic.*