Three Essays in Welfare and Health Economics: Social Choice, Health Capital, and Health Expenditure

THREE ESSAYS IN WELFARE AND HEALTH ECONOMICS: SOCIAL CHOICE, HEALTH CAPITAL, AND HEALTH EXPENDITURE

BY

JUNYING ZHAO, B.Sc. (HONORS), B.Med., M.Sc., M.P.H.

A THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2015)	McMaster University
(Health Policy PhD Program)	Hamilton, Ontario, Canada

TITLE:	Three Essays in Welfare and Health Economics: Social Choice, Health Capital, and Health Expenditure
AUTHOR:	Junying Zhao M.P.H. Harvard University, Boston, United States M.Sc. Chinese Academy of Medical Sciences, Beijing, China B.Sc. (Honors) McMaster University, Hamilton, Canada B.Med. Hebei Medical University, Shijiazhuang, China
SUPERVISOR:	Professor Jeremiah Hurley Professor Deirdre Haskell Professor William Scarth Professor Lonnie Magee
NUMBER OF PAGES:	xvii, 136

To all the people who have helped me in life

Abstract

This dissertation contains three chapters: one on social choice theory in the field of welfare economics; one that integrates health capital into endogenous growth theory in the fields of health economics and macroeconomics; and one that investigates health expenditure forecasting in the fields of health economics and economic forecasting.

The leading chapter titled **"The Possibility of Anonymous Social Orderings Using Curvature of Indifference Hypersurfaces"** is a theoretical analysis that concerns the aggregation of individual preferences into a social ordering. Working in a higher-dimensional economic environment where an indifference hypersurface is the level set of a utility function representing a preference relation, we relax the standard IIA assumption by introducing information regarding the curvature of the indifference hypersurface, which partly describes the shape of an indifference hypersurface. We show that, using curvature information, it is possible to construct a rational, anonymous social ordering function that satisfies a weaker version of IIA. The minmax-like definition of our social ordering function coincides with the Rawlsian difference principle under certain circumstances. The importance of this pure theorem is to show that the conditions for democracy can be weaker than many think.

The second chapter **"Health Capital and Endogenous Growth Theory"** is an applied theoretical paper at the intersection of macroeconomics and health. It

aims to unravel the competing effects of the health investment. Investing in health involves trade-offs such as short-term loss in consumption and the associated loss in welfare versus longer-term enhanced health, which directly improves the welfare and can potentially generate productivity gains that increase long-run consumption and welfare. This chapter examines an endogenous growth model of health investment in a two-sector economy. It explores, both analytically and numerically, the equilibrium shift and transitional dynamics after a once-for-all policy initiated by the government that reallocates labor from manufacturing to the health sector for the purpose of investing in health capital. We find that such a health investment policy improves health status in the long run, but harms the economic growth in both short and long term. The relative sizes of these competing effects depend on the specific health parameters of a country. Within the plausible range for the value of health relative to consumption, households gain welfare in the long run as long as the effectiveness of labor in health production is large. The health investment policy only makes households worse off if both labor is not productive in producing health and households value health relatively low versus consumption. For developed countries with publically-financed health sectors such as Canada where the productivity-enhancing effect of health is small, the investment in health improves welfare but harms economic growth. When these countries care welfare more than economic development, the substantial long-term gains in health and welfare are worth the short-term small pain in consumption and the long-term moderate harm in economic growth. The findings challenge the policy rationales of World Bank (1993) and World Health Organization (2001) in the sense that good health, though improves welfare, increases neither substantially the economic productivity of workers nor the economic growth rate of countries. The model itself serves as a sensitivity test for those in Van Zon and Muysken (2005); Hall and Jones (2007) in several ways. It is hoped that the relative simplicity of our model, compared to some recent models in the existing theoretical literature, can help close the gap between the formal academic work on this topic and the actual debates among policy makers in both developed and developing countries.

The third chapter **"Forecasting Health Expenditure: Methods and Applications to International Databases"** is an empirical piece in health economics. It examines a number of issues encountered when using standard health accounts data to forecast national health expenditures. In particular, it focuses on measurement issues, model specifications, and a comparison of performance indicators based on commonly used health accounts data from OECD. It assesses the performance of alternative forecasting methods based on three criteria — accuracy, precision, and certainty. Based on these criteria, it assesses the performance of model specifications including univariate (i.e., health spending) and multivariate (e.g., macroeconomic factors), static (e.g., fixed effect) and dynamic (e.g., dynamic panel), and single-equation models (e.g., ARIMA) and system of equations (e.g., VAR). It uses the better-performing models to forecast health expenditures for individual countries. This analysis makes three contributions to the literature on forecasting health expenditures. First, with longer data series, in contrast to some previous papers on health expenditure projections, we obtain a result that is more conventional in the forecasting field — econometric time series models and statistical smoothing models perform better than econometric panel data models. Second, a recent literature review of health expenditure forecasting suggests that with better computing power and more refined data, the future of forecasting is complicated micro models. But modeling and understanding the determinants of expenditure growth (whether using micro data or CGE macro models) require considerably more data and effort, and may still do worse with pure forecasting. This chapter confirms this. At the same time, it contributes to the call for more rigorous methods of forecasting, for more transparency, and for better assessment of performance. This analysis can inform both research and policy debate on budgetary planning and fiscal sustainability of health expenditure.

Acknowledgements

I am deeply grateful to my supervisor Professor Jeremiah Hurley. He was the reason I came to McMaster University to learn research on the border of health and welfare economics. He is very open-minded — teaches me the subtle balance between efficiency and equity regarding the fairness of production and allocation. Together with him, I read introductory works in social choice theory, which triggered my curiosity to dig deeper and finally led to the leading chapter of this dissertation. He is extremely supportive both financially and intellectually. He allowed me to take math courses relevant to my thesis, and finally supported me to conduct my Ph.D. thesis and pursue the honors B.Sc. Degree in math concurrently. He is incredibly caring and altruistic — protects me from all kinds of teaching/research assistance works and external projects, gives me all the time and freedom to pursue whatever inspires me at heart, which is the highest respect for a researcher. He is my role model of being a true mentor.

I am also immensely grateful to my committee member Professor Deirdre Haskell from the Math Department. She saved me when I was desperate to understand the social choice literature. She offered a reading course with me, who had no math background at that time. She taught me starting from basic logical symbols all the way through the most recent literature in this field. She required nothing but my willingness to learn. She always asks me how do I think about a subject, a definition, or some new idea to attack an unexpected problem we both are facing. My points are always seriously considered. That is the first time I understand what it means to "create" something new, rather than to "produce from the assembly line". Something that is from a new (higher dimensional geometric) point of view that the majority economists do not imagine. She introduced me to the senior mathematician in this field, and recommended to me a very unique math-oriented program to do fun, deep, and perhaps paradigm-shifting research on the border of math and economics. Because of her, I am able to embark on this challenging and rewarding intellectual journey. She is my role model of being a female mathematician/economic theorist.

I also wish to thank my committee member Professor William Scarth. He arranged my learning process of the macroeconomic modeling very logically and realistically so that I never feel worried but enjoy every step of intellectual challenge while progressing in time. He is also very knowledgeable regarding economics in general, and always has the open-door policy for me to discuss reflections about paradigm-shifting research. He shared with me a lot of helpful tips in career — his lessons in the job market decades ago, as well as his experiences to respect and communicate with seniors in the academic community. He conveyed to me his lifelong wisdom and nurtured in me a strong call to carry on the responsibility and professionalism of academia.

I would also like to thank Professor Lonnie Magee for his expertise in econometrics, patient explanation, responsive feedback, and constructive comments on my forecasting chapter.

Many people have inspired me during my writing the social choice chapter. At the Math Department, Professor McKensie Wang taught me advanced differential geometry and listed out alternative strategies for me to prove the conjecture in this chapter. Professor Ian Hambleton taught me the basic differential geometry, which helped me to establish a firm understanding of indifference surfaces in lower dimensions. Professor Andrew Nicas taught me geometric topology and discussed deeply with me the curvature of the universe though not of indifference surface in particular. Dr. Martin Bays explained to me why preference is a "relation", a more general concept than a "one-to-one" and "onto" function. Dr. Christopher Cappadocia helped me in initial thoughts of calculating the curvature of curves and always shared intriguing ideas in decision theory. At the Economics Department, Professor Seungjin Han offered helpful and general comments on the first readable draft. Dr. Wei Yang as an intelligent reader tested whether the curvature notion of indifference curves makes sense. Outside of McMaster, Professor Marc Fleurbaey at Princeton University was very responsive to explain the two proofs in his paper which my chapter extends. Professor Kam Yu at Lakehead University discussed the curvature associated with the CES utility functions in particular and urged us to verify the denominator of the curvature formula is never zero. Professor François Maniquet at the Université catholique de Louvain pointed out that the domain of social ordering function may be restricted by our two assumptions.

For the macroeconomic chapter, I thank Professor David Earn at the Math Department who taught me the qualitative theory of dynamic systems, which strengthens the mathematical foundation for my knowledge to write this chapter. For the forecasting chapter, I am truly thankful to Professor Thomas Getzen for his lengthy, detailed comments on my draft, and for his generosity with his time answering every question I had regarding forecasting. I appreciate Professor Arthur Sweetman for pointing out the two different ways to exponentiate the logarithmized data and Professor Michel Grignon for recommending me data analysis opportunity for the World Bank health unit. I acknowledge the learning experience about National Health Accounts at the Department of Health Systems Financing, World Health Organization.

Many thanks to the staff at our program — Ms. Lydia Garland, Gioia Buckley, Lyn Sauberli, and Li Wang for their care and support. I also thank Professor David Lozinski and Ms. Domenica Mazepa at the Math Department for their responsive administrative services for my cross-department course registration. I acknowledge Dr. Yang Su at the Chinese Academy of Sciences for organizing my math final exam remotely. I thank my fellow students Sandra Milicic for her warmth, support, and comradeship; Ahmad von Schlegell, Lan Wei, Daniel Patino and Kaelan Moat for positively sharing their true passions in what they do; and Moriah Magcalas who is much younger than me but earns my admiration for her math genius.

I truly appreciate my former Professors Joseph Newhouse and David Hemenway at Harvard University, and Professor Wen-Lin Chou at the Chinese University of Hong Kong for their references to McMaster University. Otherwise, I would not experience such a significant intellectual growth here. I also thank my current host program — Health Policy Ph.D. Program for designing the economics specialty and for carrying forward the spirit of interdisciplinarity.

I am eternally indebted to my parents, Zhengzhong Zhao and Yongmao Wang, for their unconditional love, unlimited support, and understanding my major shifts from medical sciences to social sciences, then to mathematical sciences. Their values on making patients and those who need help better have shaped the meaning of my life. Their emphasis on education has directed my career to academia. The Children's

Encyclopedia of Sciences they bought for me twenty years ago has ignited the fire of curiosity in my belly and lightened my circuitous way of pursuing the truth and welfare.

Contents

List of Figures

Introduction

This dissertation contains three chapters on health expenditure forecasting in the fields of health economics and economic forecasting, health capital in endogenous growth theory in the fields of health economics and macroeconomics, and social choice theory in the field of welfare economics. The health expenditure forecasting chapter empirically applies and compares forecasting methods to health expenditures. The health capital chapter theoretically examines the trade-offs between investments in the health and non-health sectors in a macroeconomic setting. The social choice chapter theoretically proves the possibility of a rational and democratic (non-dictatorial) social aggregation of free individual preferences from a geometric point of view.

These three chapters differ not only in fields but also in the ways of thinking from empirical, to applied theoretical, then to pure theoretical, which is the order in which each chapter was conducted, and the order that reflects my intellectual growth throughout my pursuit of this Ph.D. Below I introduce the three chapters in this order. But in the body of the dissertation, I arrange these three chapters in the order of my preferred way of thinking from pure theoretical, to applied theoretical, then to empirical.

Health expenditures have been increasing in the past decades among both developed and developing countries (Gerdtham *et al.*, 1992; Lorenzoni *et al.*, 2014).

This raises global concerns about fiscal sustainability and calls for policy interventions (European Union, 2009; Centers for Medicare & Medicaid Services, 2011). This chapter primarily concerns a methodological issue — how can we best forecast health expenditure in a systematic way; and then an empirical question — given our best estimates, how much would the worldwide health expenditures be in the short- and medium-run future? This chapter applies a comprehensive set of measurements and projection models for forecasting, carefully tests their performances based on formal criteria, and uses recent and comparable data provided by international sources. The methods and empirical results of health expenditure forecasts can inform policy makers regarding budget planning, and financial gaps between the amounts of monetary resources needed and those available.

The chapter draws on two sub-fields of economics: health economics (esp., expenditure), and economic forecasting. The majority of the health expenditure literature tries primarily to understand past drivers of the health expenditure, rather than to project health spending into the future. Few studies integrate health expenditure estimation and economic forecasting. A seminal paper in the literature on forecasting health expenditures, Getzen and Poullier (1992), obtains an unusual result — that panel data models performed better than time-series models — and argue that people should examine this issue as more data accumulate. This chapter does exactly this. Exploiting a much longer times -series, we obtain the more conventional result, in a sense, overturning their initial finding. Second, Astolfi *et al.* (2012) conduct a comprehensive literature review of health expenditure forecasting and suggest that with better computing power and more refined data, the future of forecasting is complicated micro-level models. However, readers should carefully distinguish the objective of simply getting accurate future forecasts of spending from modeling and understanding past determinants of expenditure. Modeling and understanding the determinants of expenditure growth (whether using micro data or CGE macro models) require considerably more data and effort, and may still do worse for pure forecasting. This chapter confirms this. At the same time, this chapter contributes to the call (articulated in Astolfi *et al.* (2012)) for more rigorous methods, for more transparency, and for better assessment of performance.

During my work on the forecasting chapter, I came to realize that the reduced-form models fail to specify underlying causal mechanisms and are narrow within the health sector. These shortcomings of the empirical approach spurred me to know more about "why and how", and to know more beyond but relevant to the health sector from a broader view of the entire economy. Also, in reality, investing in the health sector has been promoted worldwide for economic development (World Bank, 1993; World Health Organization, 2001). Yet, investing in the health sector involves trade-offs whose net effect is still ambiguous. So I started to conduct the macroeconomic theory chapter to disentangle these competing effects of investing in health capital.

The most direct effect of such an investment is positive: it makes people healthier. But there is also an immediate and permanent negative effect since such an investment draws resources from the non-health sector, causing households to have fewer consumption goods, and this harms their well-being in both the short and the long run. Meanwhile, workers that remain in the manufacturing sector are now healthier, and, therefore, more productive. This indirect effect has two implications. First, it tempers and could even reverse the reduction in manufacturing output. Second, it raises the marginal product of physical capital, hence stimulates investment in capital accumulation. As a consequence, if healthier workers were sufficiently more productive, there would be a win-win outcome — more health and more economic growth in the long term. Otherwise, this win-win outcome would not occur.

This chapter is not the first attempt to unravel these competing effects. But macro-theoretical studies in health economics are relatively few. The available theoretical modelings disentangling such trade-offs are incomplete especially those focusing on a publically financed health sector. Therefore, however tempting the policy prescriptions of health investment may sound, it is concerning that they are based on such a limited analytical support. From a macroeconomic policy point of view, an extended growth model is needed that is both rigorous in specifying the mechanisms whereby health capital affects utility and productivity and accessible to policy makers. This chapter uses a relatively simple endogenous-growth macroeconomic model to assess the relative importance of these competing effects of investing in health. It makes several contributions to the literature. It disentangles the theoretical ambiguities of the trade-offs of introducing a health investment policy — in this case modeled as a reallocation of labor from the manufacturing to the health sector. The model is consistent with the modern macroeconomic requirement of internal consistency and optimization on the part of economic agents, and yet is more accessible to policy makers than much existing work in this area. The model itself serves as a sensitivity test for those in Van Zon and Muysken (2005); Hall and Jones (2007). It is hoped that the simplicity of our model and its focus on publicly financed health investment contribute to the policy debate in countries with a publically funded health sector such as Canada. Alternative calibrations of the model further allow us to address the challenges faced by both developed and developing countries with publicly-financed health sectors.

Again, during my work on the health capital chapter, although I agreed with the famous "micro-founded macro" Lucas Critique, I came to doubt the summing of individual utilities into the societal utility, which is a conventional form of the social welfare function. Inspired by this doubt, and my long-standing interest in welfare economics, I started to read and ask research questions in social choice theory.

The Nobel Laureate K. Arrow "introduced a general approach to the study of preference aggregation, partly inspired by his teacher of logic, Alfred Tarski, from whom he had learned relation theory." (Stanford Encyclopedia of Philosophy, 2013) His striking Impossibility Theorem is commonly acknowledged as the basis of the modern social choice theory. This theorem says that there does not exist a rational social ordering function simultaneously satisfying Unrestricted Domain, Pareto, Independence of Irrelevant Alternatives (IIA), and Non-dictatorship (Sen, 1986; Mas-Colell *et al.*, 1995). Such non-existence mainly comes from the strong condition of IIA, which requires the social ordering between any pair of alternatives to be aggregated from individual rankings over that pair only. To obtain an existence result, the literature has been relaxing this condition. Inada (1964) and Mayston (1980) define the marginal rate of substitution (MRS) and show that the additional requirement of each individual's MRSs being equal in small neighborhoods under any two preference profiles still leads to impossibility results. More recently, Fleurbaey *et al.* (2005) weaken the IIA condition by adding information about indifference surfaces to the traditional condition. The additional information in a successfully proved proposition is a number assigned to the point of intersection of each indifference surface and a monotone path from the origin that contains relevant benchmark bundles. They then use the same path to construct an anonymous social ordering function that satisfies this weakened IIA.

This chapter responds to the call of Fleurbaey *et al.* (2005) to investigate how other types of information about indifference surfaces may help the construction of non-dictatorial social ordering functions that satisfy IIA conditions weakened in various ways. In particular, we introduce information about the curvature of the indifference hypersurfaces associated with the allocations under consideration. Curvature depends on the second derivative of an indifference hypersurface, that is, it is a measure of the rate of change in the MRS among goods. The notion of curvature was introduced to economics by another Nobel Laureate, Debreu (1972), and more recently has been used by Hayashi (2008) in the theory of general equilibrium. We make two very weak assumptions about the curvature of an indifference hypersurface (and conjecture that, for every smooth preference relation that satisfies monotonicity and continuity, these two assumptions are satisfied.) Under these assumptions, we define an alternative way to relax IIA using the information on curvature. This IIA requires less information about the indifference hypersurface than the IIA in Fleurbaey *et al.* (2005, Proposition 5). We then show that, by using the information on the curvature of indifference hypersurfaces one can construct a social ordering function that satisfies the corresponding weaker version of IIA and anonymity. Such a result is as general as Fleurbaey *et al.* (2005, Proposition 5). It also extends Inada (1964) and Mayston (1980) by showing that, in addition to the first derivative notion, i.e., MRS, the second derivative notion of the indifference hypersurface can lead to possibility results. Further, we show that such a curvature-based social ordering function has some correspondence with the widely analyzed Rawlsian maximin social choice rule.

Chapter 1

The Possibility of Anonymous Social Orderings Using Curvature of Indifference Hypersurfaces

1.1 Introduction

Arrow's Impossibility Theorem shows that there does not exist a rational social ordering function simultaneously satisfying Unrestricted Domain (UD), Pareto (P), Independence of Irrelevant Alternatives (IIA), and Non-dictatorship (ND) (Sen, 1986; Mas-Colell *et al.*, 1995). Such non-existence mainly comes from the strong assumption of IIA, which requires the social ordering between any pair of alternatives to be aggregated from individual rankings over that pair only. In recent literature, two different approaches are taken to relax these conditions. One approach, adopted in abstract, non-economic environments, has been to relax the UD condition either by restricting the domain of alternatives or by restricting the domain of preferences over alternatives of a social ordering function. Kalai *et al.* (1979) restricts the preference domain to be *saturating*, i.e., there are at least two nontrivial pairs and any two nontrivial pairs are connected. Le Breton and Weymark (1996) further requires all nontrivial pairs to be connected. Both show that a social ordering function that satisfies restricted domain and the other two conditions is still dictatorial. Redekop (1991) restricts the preference domain to be topologically small so that it does not display much diversity in individual preferences. In this situation, he shows that if the preference domain is topologically large, then there does not exist a non-dictatorial social ordering function. Campbell and Kelly (2007, 2009) restrict the domain of infinite alternatives to a domain with a sufficient set of alternatives as its proper subset, which leads to a non-dictatorial (though sub-dictatorial) social ordering function.

Another approach, often adopted in an economic environment, has been to relax the IIA condition. Inada (1964) and Mayston (1980) define the marginal rate of substitution (MRS) and show that the additional requirement of each individual's

MRSs being equal in small neighborhoods under any two preference profiles still leads to impossibility results. More fundamentally, Campbell (1992, Theorem 10.13) shows that if preferences are continuous, then a social ordering function satisfies IIA if and only if it is constant or dictatorial. More recently, Fleurbaey *et al.* (2005) weaken the IIA condition by adding information about indifference surfaces to the traditional assumption. In their Proposition 3, for example, this information takes the form of the requirement that each individual's indifference surfaces within a neighborhood around the consumption bundles are identical under the two profiles. They define a social ordering function which satisfies this weakened IIA and is non-dictatorial, though sub-dictatorial. In their Proposition 5, following ideas from Pazner and Schmeidler (1978), the additional information is a number assigned to the point of intersection of each indifference surface and a monotone path from the origin that contains relevant benchmark bundles. They then use the same path to construct an anonymous social ordering function that satisfies this weakened IIA.

This paper responds to the call of Fleurbaey *et al.* (2005) to investigate how other types of information about indifference surfaces may support the construction of non-dictatorial social ordering functions that satisfy IIA conditions weakened in various ways. In particular, we introduce information about the curvature of the indifference hypersurfaces associated with the allocations under consideration. Curvature depends on the second derivative of an indifference hypersurface, that is, it is a measure of the rate of change in the MRS among goods. The notion of curvature is introduced by Debreu (1972) to economic literature and more recently has been used by Hayashi (2008) in the theory of general equilibrium. We make two very weak assumptions about the curvature of an indifference hypersurface (and conjecture that, for every smooth preference relation that satisfies monotonicity and continuity, these assumptions are satisfied.) Under these assumptions, we define an alternative way to relax IIA using the information of curvature. This IIA requires less information about the indifference hypersurface than the IIA in Fleurbaey *et al.* (2005, Proposition 5). We then show that, by adding information on the curvature of indifference hypersurfaces one can construct a social ordering function that satisfies the corresponding weaker version of IIA and anonymity. Such a result is as general as Fleurbaey *et al.* (2005, Proposition 5). It also extends Inada (1964) and Mayston (1980) by showing that, in addition to the first derivative notion, i.e., MRS, the second derivative notion of the indifference hypersurface can lead to possibility results. Further, we show that such a curvature-based social ordering function has some correspondence with the widely analyzed Rawlsian maximin social choice rule.

1.2 Curvature of Indifference Hypersurface

Fix a set of individuals Ind = {1, ..., n } and a set of goods G = {1, ..., ℓ }. An allocation is a matrix $\mathbf{x} \in \mathbb{R}_+^{n\ell}$, in which the n rows are vectors of length ℓ where for each i, $x_i = (x_{i1}, \cdots, x_{i\ell})$ is the consumption bundle of individual i, and the ℓ columns are vectors of length n where for each k, $x_k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix}$ represents the distribution of the k-th good amongst the n individuals. We assume that each individual i has a weak preference relation R_i (resp. strict preference P_i), that is rational, i.e., complete ($\forall x, y \in \mathbb{R}_+^\ell$, $\neg(xRy)$ implies yRx) and transitive ($\forall x, y \in \mathbb{R}_+^\ell$, xRy and yRz imply xRz), and that further satisfies montonicity ($x \gg y$ implies xRy), convexity (yRx

and zRx imply $\alpha y + (1-\alpha)zRx$ for any $\alpha \in [0,1]$) and continuity (R is preserved under limits). Any such preference relation can be represented by a continuous real-valued utility function; that is, for every R there exists u such that xRy if and only if $u(x) \geq u(y)$ (Mas-Colell *et al.*, 1995).

Definition 1. For any $z \in \mathbb{R}^{\ell}_{+}$ and for any preference relation R, the indifference set through z with respect to R is

$$I(z, R) := \{w \in \mathbb{R}^{\ell}_{+} : zRw \ \& \ wRz\}.$$

Equivalently, if u is any utility function which represents the preference relation, the indifference set is a level hypersurface of u; that is,

$$I(z, R) := u^{-1}(k) = \{w \in \mathbb{R}^{\ell}_{+} : u(w) = k\}.$$

Notice that $\{(w, k) : u(w) = k\}$ is an ℓ-dimensional hypersurface in $\mathbb{R}^{\ell+1}_{+}$ space. Any level set is an $\mathbb{R}^{\ell-1}_{+}$-dimensional set. We further assume that each R_i has the property that its indifference hypersurfaces are at least twice continuously differentiable, and hence have a well-defined curvature. We define curvature formally below, but here we first motivate why information on curvature may be relevant to the construction of a social ordering function.

Let us consider a consumption bundles z and two preference relations R_1, R_2. Suppose that the two hypersurfaces $I(z, R_1)$, $I(z, R_2)$ have the same slope at z, i.e., same MRS at z, but different second derivatives at z such that the curvature of $I(z, R_2)$ is larger than the curvature of $I(z, R_1)$ (see Figure 1.1). In a ball of sufficiently small radius ϵ centered at z, the fact that the curvature of $I(z, R_2)$ is greater than the

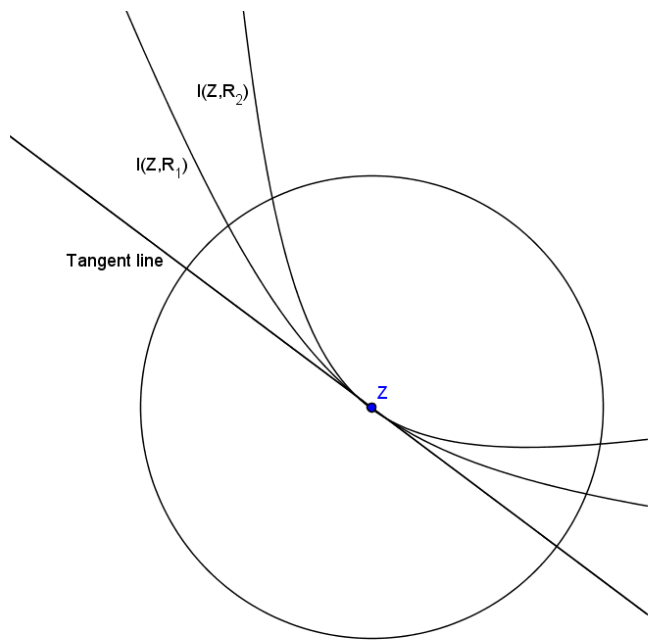


Figure 1.1: Indifference curves with the same first derivative but different curvatures at the center z in a neighborhood.

curvature of $I(z, R_1)$ implies that the area enclosed by the ball and below $I(z, R_2)$ is greater than the area enclosed by the ball and below $I(z, R_1)$. That is, the larger curvature corresponds to a larger lower contour set in the neighborhood. Because of the curvature of $I(z, R_2)$ is greater than the curvature of $I(z, R_1)$, in the following sense the preference for z can be said to be stronger under R_2 than under R_1: under R_2 the bundle z is preferred to a larger set of alternative bundles within this standardized neighbourhood.

We generalize this idea of curvature to higher dimensions using the Gauss-Kronecker curvature (Thorpe, 1979, Theorem 12.5). The Gauss-Kronecker curvature at a point w in any $(\ell-1)$-hypersurface is defined as follows.

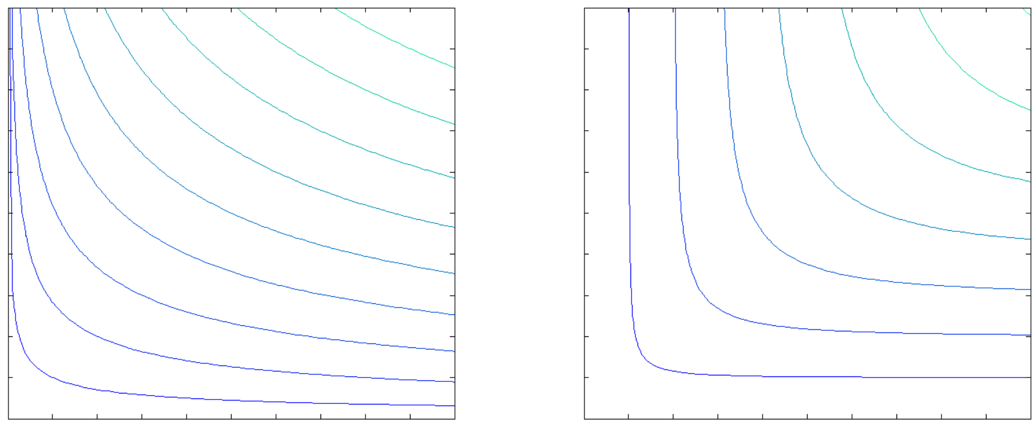
Definition 2. Let S be an oriented $(\ell-1)$-hypersurface in $\mathbb{R}^{\ell+1}_{+}$ and let $w \in S$. Let $\mathbf{N}$ be any non-zero normal vector field on S and let $\{v_1, \ldots, v_{\ell-1}\}$ be any basis for the vector field S_w tangent to S at w. The absolute value of the Gauss-Kronecker curvature of S through z at the point w is

$$c(S,w) := \left| \det \begin{pmatrix} \nabla_{v_1}\mathbf{N} \\ \vdots \\ \nabla_{v_{\ell-1}}\mathbf{N} \\ \mathbf{N}(w) \end{pmatrix} \Big/ (-1)^{\ell-1} \|\mathbf{N}(w)\|^{\ell-1} \det \begin{pmatrix} v_1 \\ \vdots \\ v_{\ell-1} \\ \mathbf{N}(w) \end{pmatrix} \right|.$$

In this definition, the numerator is the determinant of the bordered Hessian matrix, $\det(B_\ell(w))$, of the function whose level set is S. Notice that $B_\ell(w)$ is constructed by putting a border consisting of the first partial derivatives of the function with respect to each coordinate into the matrix of the second partial derivatives (Sydsæter *et al.*, 2005, p.74). The terms in the denominator only include the first partial derivatives. Since Hayashi (2008, p.363), Dierker (1975, p.49), and Debreu (1972, Equation 3) are only concerned with whether the Gauss-Kronecker curvature is non-zero, they omit the denominator and define the curvature as the determinant of the bordered Hessian.

The formula applies in particular when $S = I(z, R)$ is an indifference hypersurface. We illustrate relevant properties of curvature using indifference hypersurfaces derived from commonly used functions. For example, consider a typical utility function — Cobb-Douglas. Notice that a unique preference relation, associated with a unique indifference hypersurface, can be represented by different utility functions. That is, the curvature value of an indifference hypersurface does not depend on a particular utility function.

Given a 2-variable specification $x^{1/3}y^{2/3}$, the Gauss-Kronecker curvature of the



(a) Cobb-Douglas (b) Constant Elasticity of Substitution

Figure 1.2: Indifference sets as level sets of typical utility functions, demonstrating the behavior of the curvature.

corresponding indifference hypersurface is

$$c(S, x) = \frac{6k^{3/2}x^2}{(k^3 + 4x^3)^{3/2}}.$$

For any fixed utility level k, $\lim_{x\to\infty} c(x) = 0$ and $\lim_{x\to 0} c(x) = 0$. As $c(x)$ is defined for all $x \in (0, \infty)$, if we restrict x to lie on any given indifference hypersurface, the curvature will have a supremum value, which is achieved for some x in $(0, \infty)$. Now consider the family of indifference hypersurfaces by varying k. If we move out along any line $y = \alpha x, \alpha \in \mathbb{R}_+$, we can rewrite

$$c(S, x) = \frac{6\alpha^3 x^{13/2}}{(\alpha^6 x^9 + 4x^3)^{3/2}}.$$

For fixed α, $\lim_{x\to\infty} c(x) = 0$. Thus, along the fixed line, the curvature decreases as one moves away from the origin (Figure 1.2).

Further, given a 3-variable specification $k = x^{1/3}y^{1/3}z^{1/3}$, we derive the curvature to be

$$c(S, x, y) = \frac{3x^4y^4}{2(x^6y^4 + x^4y^6 + k^6x^2y^2) + k^6(x^4 + y^4 + k^{-12}x^8y^8)}.$$

For fixed k, the function $c(S, x, y)$ has a global maximum at $(x, y) = (1, 1)$, which means that there exists a supremum of curvatures for each indifference hypersurface. Then if the hypersurface moves out along the directions $y = \alpha x$, $z = \beta x + \gamma y$, where $\beta, \gamma \in \mathbb{R}_+$, we derive

$$c(S, x) = \eta \frac{1}{x^2},$$

where η is constructed by α, β, and γ. Fix α, β, γ, $\lim_{x\to\infty} c(x) = 0$. This means that curvature decreases as indifference hypersurface moves away from the origin.

More generally, given ℓ-variable specification $u(x_1, \ldots, x_\ell) = Ax_1^{a_1} \ldots x_\ell^{a_\ell}$, assuming $A, a_1, \ldots, a_\ell > 0$ and $a_1 + \cdots + a_\ell \leq 1$, we derive (see Appendix)

$$c(S, x_1, \ldots, x_\ell) = A^{\ell-2}\left(\prod_{i=1}^{\ell} a_i\right)\left(\sum_{i=1}^{\ell} a_i\right)\frac{\left(\prod_{i=1}^{\ell} x_i^{(\ell-2)a_i+1}\right)}{\left(\sum_{i=1}^{\ell} a_i^2 \prod_{j\neq i}^{\ell} x_j^2\right)^{3/2}}.$$

Again, for any i, as $x_i \to \infty$ and all $x_{j\neq i} \to 0$, $c(S, x_1, \ldots, x_\ell) \to 0$. This means the existence of a supremum of curvatures for each indifference hypersurface. Now fix a linear relationship among the other coordinates, and express it in terms of x_1, then $\lim_{x_1\to\infty} c(x_1) = 0$ for all $\ell \geq 2$ as desired. Such properties of curvature also hold for many other indifference hypersurfaces which are level sets of utility functions such as those with constant elasticity of substitution (CES, Figure 1.2).

In order to use curvature to define our social ordering function, we need to make two further assumptions on the individual preference relations.

Assumption 1. *For each indifference hypersurface $I(z, R)$, the supremum of the curvatures along the indifference hypersurface exists. That is,* $\sup\{c(I(z, R), w) : w \in I(z, R)\}$ *exists.*

Given Assumption 1, we can make the following definition.

Definition 3. The *Curvature of the indifference hypersurface through the point z* is a function that assigns to $I(z, R)$ the supremum of the curvatures at all points w in $I(z, R)$; that is,

$$C(z, R) := \sup\{c(I(z, R), w) : w \in I(z, R)\}.$$

Assumption 2. *For a fixed preference relation R, if $w_1 \gg w_2$, then the property $C(w_1, R) \leq C(w_2, R)$ holds. That is,* $\sup\{c(I(w_1, R), w) : w \in I(w_1, R)\} \leq \sup\{c(I(w_2, R), w) : w \in I(w_2, R)\}$.

The above examples show that these assumptions hold for indifference hypersurfaces derived from two families of typical utility functions. In fact, we conjecture that, for every smooth preference relation that satisfies monotonicity and continuity, these assumptions are satisfied.

We emphasize that utility functions discussed here are purely for the purpose of examples. The social ordering function that we will construct is defined using the curvature of indifference hypersurfaces of individuals' preference relations. It does not depend on the choice of a particular utility function that represents the preference relation.

We then investigate how C can be useful in constructing a social ordering function.

1.3 Social Ordering Function

The input of a social ordering function is an n-tuple, called a profile, of individual preference relations, $\mathbf{R} = (R_1, \ldots, R_n)$ (resp. $\mathbf{P} = (P_1, \ldots, P_n)$). Just as an individual preference is a relation between vectors of allocations, so a social preference relation is a relation between matrices of allocations representing the vectors of allocations to the n individuals in a population. A social ordering function is a function f from the set of n-tuple profiles $\mathcal{R}$ to the set of social preference relations $\mathcal{S}$; that is,

$$\begin{aligned} f :& \mathcal{R} \to \mathcal{S} \\ & \mathbf{R} \mapsto f(\mathbf{R}) \\ & \mathbf{P} \mapsto f(\mathbf{P}) \end{aligned}$$

The following properties have been central to the analysis of social ordering functions.

Pareto: f is *Paretian* if for every profile and for every pair of allocations $\mathbf{x}$ and $\mathbf{y}$, if every individual prefers her/his consumption bundle under $\mathbf{x}$ over that under $\mathbf{y}$ then allocation $\mathbf{x}$ is socially preferred over $\mathbf{y}$; that is,

$$\forall \mathbf{R} \in \mathcal{R} \ \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n\ell} \ \ \big(\forall i \in \text{Ind} \ \ x_i R_i y_i \Longrightarrow \mathbf{x} f(\mathbf{R}) \mathbf{y}\big).$$

Dictatorship: f is *dictatorial* if there exists an individual such that for every profile and for every pair of allocations $\mathbf{x}$ and $\mathbf{y}$, if (s)he strongly prefers her/his consumption bundle under $\mathbf{x}$ over that under $\mathbf{y}$, then $\mathbf{x}$ is socially strongly preferred over $\mathbf{y}$; that is,

$$\exists\, i_0 \in \text{Ind} \ \ \big(\forall \mathbf{R} \in \mathcal{R} \ \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n\ell} \ \ x_{i_0} P_{i_0} y_{i_0} \Longrightarrow \mathbf{x} f(\mathbf{P}) \mathbf{y}\big).$$

Anonymity: f is *anonymous* if for every profile and for every pair of allocations $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}$ is socially preferred over $\mathbf{y}$ if and only if $\mathbf{x}$ is socially preferred over $\mathbf{y}$ after any permutation of n individuals; that is,

$$\forall \mathbf{R} \in \mathcal{R} \ \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n\ell} \ \ \big(\mathbf{x} f(\mathbf{R}) \mathbf{y} \iff \pi(\mathbf{x}) f(\pi(\mathbf{R})) \pi(\mathbf{y})\big),$$

where π is a permutation of Ind which induces a permutation of the profile: $\pi(\mathbf{R}) = \pi(R_1, \ldots, R_n) = (R_{\pi_{(1)}}, \ldots, R_{\pi_{(n)}})$.

IIA: For every two profiles and for every pair of allocations $\mathbf{x}$ and $\mathbf{y}$, if every individual preference relation between the two consumption bundles under allocations $\mathbf{x}$ and $\mathbf{y}$ agrees under the two profiles, then the social preference relation between $\mathbf{x}$ and $\mathbf{y}$ agrees under the two profiles; that is, $\forall \mathbf{R}, \mathbf{R}' \in \mathcal{R} \ \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n\ell}$

$$\Big(\forall i \in \text{Ind} \ \ \big(x_i R_i y_i \iff x_i R_i' y_i\big) \Longrightarrow \big(\mathbf{x} f(\mathbf{R}) \mathbf{y} \iff \mathbf{x} f(\mathbf{R}') \mathbf{y}\big)\Big).$$

Arrow's Theorem tells us that the IIA condition is too strong to enable the possibility of a non-dictatorial f. Fleurbaey *et al.* (2005) has observed various alternatives for relaxing the IIA condition. Here we propose another one. We weaken it by adding some information about each indifference hypersurface which comes from the C function.

Definition 4. IIA-curvature: $\forall \mathbf{R}, \mathbf{R}' \in \mathcal{R} \ \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n\ell}$

$$\begin{aligned}\Big(\forall i \in \text{Ind } \big(x_i R_i y_i \iff x_i R_i' y_i \ \wedge \\ C(x_i, R_i) = C(x_i, R_i') \ \wedge \\ C(y_i, R_i) = C(y_i, R_i') \ \big) \\ \implies \big(\mathbf{x} f(\mathbf{R}) \mathbf{y} \iff \mathbf{x} f(\mathbf{R}') \mathbf{y}\big)\Big).\end{aligned}$$

That is, for every two profiles and for every pair of allocations $\mathbf{x}$ and $\mathbf{y}$, the social preference relation between $\mathbf{x}$ and $\mathbf{y}$ agrees under the two profiles if the following conditions hold: if every individual preference relation between the two consumption bundles under $\mathbf{x}$ and $\mathbf{y}$ agrees under every two profiles; and furthermore the supremum of the curvatures of every individual indifference hypersurface under profile $\mathbf{R}$ through consumption bundle under allocation $\mathbf{x}$ is equal to the supremum of the curvatures of every individual indifference hypersurface under profile $\mathbf{R}'$ through the same consumption bundle; and the supremum of the curvatures of every individual indifference hypersurface under profile $\mathbf{R}$ through consumption bundle under allocation $\mathbf{y}$ is equal to the supremum of the curvatures of every individual hypersurface under profile $\mathbf{R}'$ through the same consumption bundle. Notice the difference between this IIA and the IIA-ISPω_0 (Fleurbaey *et al.*, 2005, Proposition 5): any way of weakening IIA requires additional information about individual preferences under the two profiles. In IIA-ISPω_0, that additional information comes from the intersections of two indifference hypersurfaces with a chosen path. In IIA-curvature, that additional information comes from looking at the suprema of curvatures of the indifference hypersurfaces, which is more invariant.

Theorem 1. *Let $\mathcal{R}$ be the set of profiles of preference relations satisfying Assumptions*

1 and 2. Then there exists a rational social ordering function f satisfying Pareto, Non-dictatorship, Anonymity, and IIA-curvature.

Proof. Define a social ordering function as follows

$$\mathbf{x}f(\mathbf{R})\mathbf{y} \iff \max_i\{C(x_i, R_i)\} \leq \max_i\{C(y_i, R_i)\}.$$

We use the information on curvatures in two ways in the proposed social ordering function. First, drawing on the notion of curvature as representing strength of preference in the sense noted above (Figure 1.1), we assign to each allocation the maximum of the *Curvatures of the indifference hypersurfaces* through the individual/bundle pairs of that allocation. Call this curvature information, which reflects both the bundle received and the preference of the individual who receives it, the curvature index for an allocation. Second, having used the curvature index to define the social ordering function, it follows from Assumption 2 and the monotonicity of preference that the socially preferred allocation guarantees that at least one of the index individuals is better off. Notice that the two individuals at which the maxima occur on either side of the inequality are expected to be different. Hence, there are no sub-dictators.

First, we need to verify that $f(\mathbf{R})$ is indeed a complete and transitive relation. To show $f(\mathbf{R})$ is complete, we need to show that the negation of output $\mathbf{x}f(\mathbf{R})\mathbf{y}$ gives $\mathbf{y}f(\mathbf{R})\mathbf{x}$. That is, to show $\neg\ (\mathbf{x}f(\mathbf{R})\mathbf{y}) \Rightarrow \mathbf{y}f(\mathbf{R})\mathbf{x}$. If $\neg\ (\mathbf{x}f(\mathbf{R})\mathbf{y})$, then $\max_i\{C(x_i, R_i)\} > \max_i\{C(y_i, R_i)\}$, and hence $\mathbf{y}f(\mathbf{R})\mathbf{x}$. So our f is complete, as required.

To show $f(\mathbf{R})$ is transitive, we need to show that if allocation matrix $\mathbf{x}$ is socially preferred over $\mathbf{y}$, and $\mathbf{y}$ over $\mathbf{z}$, then $\mathbf{x}$ is socially preferred over $\mathbf{z}$. Given that $\mathbf{x}f(\mathbf{R})\mathbf{y}$

if and only if $\max_i\{C(x_i, R_i)\} \leq \max_i\{C(y_i, R_i)\}$, and that $\mathbf{y}f(\mathbf{R})\mathbf{z}$ if and only if $\max_i\{C(y_i, R_i)\} \leq \max_i\{C(z_i, R_i)\}$, if $\mathbf{x}f(\mathbf{R})\mathbf{y}$ and $\mathbf{y}f(\mathbf{R})\mathbf{z}$, then

$$\max_i\{C(x_i, R_i)\} \leq \max_i\{C(y_i, R_i)\} \leq \max_i\{C(z_i, R_i)\}.$$

The inequalities always hold for these three maxima no matter how i differs. So $\max_i\{C(x_i, R_i)\} \leq \max_i\{C(z_i, R_i)\}$, and hence $\mathbf{x}f(\mathbf{R})\mathbf{z}$, as required.

To show that f is Paretian, we need to verify that if $x_i R_i y_i$ for every i, then $\mathbf{x}f(\mathbf{R})\mathbf{y}$. So assume $x_i R_i y_i$ for every i. For each i, fix any vector $\vec{v}_i$ through the origin, and let w_i, z_i be the points where $\vec{v}_i$ intersects $I(x_i, R_i)$ and $I(y_i, R_i)$ respectively. As $x_i R_i y_i$, it follows by transitivity that $w_i R_i z_i$ and hence by monotonicity $w_i \gg z_i$ (as w_i, z_i lie on a vector through the origin, either $w_i \gg z_i$ or $z_i \gg w_i$). By Assumption 2, $C(w_i, R_i) \leq C(z_i, R_i)$ and hence $C(x_i, R_i) \leq C(y_i, R_i)$. Now suppose $\max_i\{C(x_i, R_i)\} = C(x_{i_1}, R_{i_1})$ and $\max_i\{C(y_i, R_i)\} = C(y_{i_2}, R_{i_2})$. Then $C(x_{i_1}, R_{i_1}) \leq C(y_{i_1}, R_{i_1}) \leq C(y_{i_2}, R_{i_2})$ and hence $\mathbf{x}f(\mathbf{R})\mathbf{y}$, as required.

No matter how we permute $i \in \text{Ind}$, the finite set $\{c(x_i, R_i)\}$ is permuted but otherwise unchanged, so the $\max_i\{C(x_i, R_i)\}$ is unchanged. Similarly, the $\max_i\{C(y_i, R_i)\}$ is unchanged. Hence the resulting $f(\mathbf{R})$ regarding any pair $\{\mathbf{x}, \mathbf{y}\}$ remains unchanged after permutations. Our f is anonymous, as required.

Finally, because f is defined using $C(x_i, R_i)$ and the assumption of $IIA{-}curvature$ is that the values of C remain unchanged from $\mathbf{R}$ to $\mathbf{R}'$, also the output of f will remain unchanged. $\square$

1.4 Generalization

The proof that the social ordering function defined above satisfies all the desired properties depends only on the fact that $C(x, R)$ satisfies Assumptions 1 and 2 about curvature. Instead of the curvature function C from the set of indifference hypersurfaces to $\mathbb{R}_+$, we could use any function C that has the following monotonicity property: if $w \gg z$ then $C(w, R) \leq C(z, R)$. The same definition will give a social ordering function that is rational and anonymous, and satisfies Pareto and IIA-curvature.

Furthermore, we could also use any function C which satisfies the reverse inequality: $w \gg z$ implies $C(w, R) \geq C(z, R)$. Defining f by $\mathbf{x} f(\mathbf{R})$ if and only if $\min_i\{C(x_i, R_i)\} \geq \min_i\{C(y_i, R_i)\}$ gives a social ordering function which is rational, anonymous, and satisfies IIA-curvature by the same argument with inequality reversed. The definition of Fleurbaey *et al.* (2005, Proposition 5) falls under this description.

1.5 Curvature and the Rawlsian Difference Principle

Our curvature-based social ordering function turns out to have some correspondance with the Rawlsian difference principle. Consider any pair of allocation matrices $\{\mathbf{x}, \mathbf{y}\}$. Let $A(\mathbf{x})$ be $\{C(x_1, R_1), ..., C(x_n, R_n)\}$, which is the set of all suprema of curvatures at all points in indifference hypersurfaces for all $i \in \text{Ind}$. The intuition of our social ordering function is first to identify the individuals with the largest curvatures of indifference hypersurfaces. So choose the maximum of A for allocation $\mathbf{x}$, denoted as $\max A(\mathbf{x})$. Similarly for allocation $\mathbf{y}$, we choose $\max A(\mathbf{y})$. Then choose the minimum

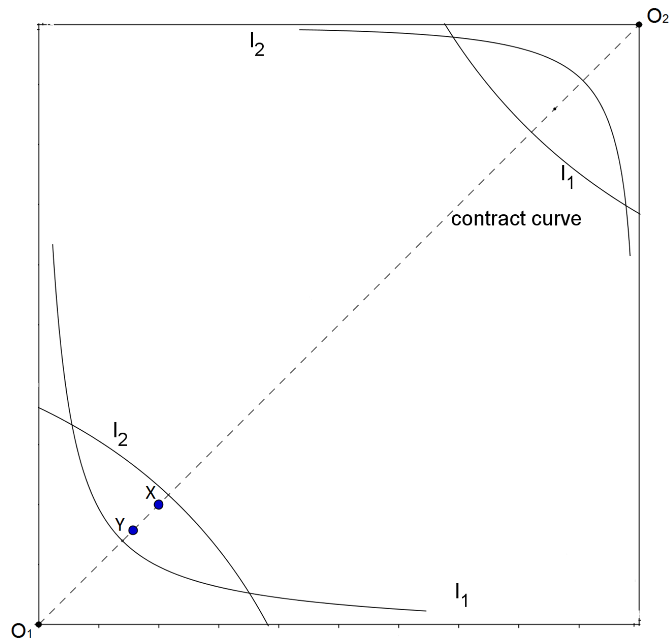


Figure 1.3: The Edgeworth box of two-individual, two-good economy where the more equal allocation $\mathbf{x}$ is socially preferred over $\mathbf{y}$ on the contract curve by the curvature-based social ordering function.

from the set $\{\max A(\mathbf{x}), \max A(\mathbf{y})\}$. Our social ordering function is equivalent to the definition as follows: $\mathbf{x}f(\mathbf{R})\mathbf{y}$ if and only if $\min\{\max A(\mathbf{x}), \max A(\mathbf{y})\} = \max A(\mathbf{x})$. Under certain conditions, this social ordering function assigns social preference to allocation $\mathbf{x}$ over $\mathbf{y}$ if and only if the most disadvantaged benefits more under $\mathbf{x}$ than $\mathbf{y}$. In this sense, it corresponds to the Rawlsian difference principle of preferring the allocation that benefits the most disadvantaged individual.

To see this, consider a simple two-person, two-good exchange economy represented in an Edgeworth box. Consider two Pareto efficient allocation matrices on the contract curve, $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}$ is located closer to O_1 than to O_2, and $\mathbf{y}$ is still closer than $\mathbf{x}$ to O_1. Assume in particular that the two agents have identical preferences symmetric about the contract curve (Figure 1.3). Identifying the preferred allocation proceeds in three steps. Step 1: because $C(x_2, R_2) \leq C(x_1, R_1)$, choose $C(x_1, R_1)$ under $\mathbf{x}$. Step 2: because $C(y_2, R_2) \leq C(y_1, R_1)$, choose $C(y_1, R_1)$ under $\mathbf{y}$. Step 3: because $C(x_1, R_1) \leq C(y_1, R_1)$, society prefers $\mathbf{x}$ over $\mathbf{y}$, which corresponds to the Rawlsian difference principle. Indeed, in this particular case, the social ordering function will always choose the allocation that is closest to an equal division of the goods, which corresponds to the Rawlsian difference principle.

In a more general case without assuming symmetry of indifference sets, this correspondence still holds. The reader can easily see this by orientating individual two's coordinate system 180-degree clockwise, equating the two origins, applying Assumption 2 and identical agents assumption while proceeding through the three steps to determine the social ordering.

Even more generally, this result continues to hold in certain cases when relaxing the assumption of identical preferences. Without such an assumption, there are four cases to consider. Case 1: $C(x_2, R_2) \leq C(x_1, R_1)$ and $C(y_2, R_2) \leq C(y_1, R_1)$. The choice made by our social ordering function corresponds to that made by the Rawlsian difference principle. Case 2: $C(x_2, R_2) \leq C(x_1, R_1)$ and $C(y_1, R_1) \leq C(y_2, R_2)$. Because $C(y_1, R_1) \geq C(x_1, R_1)$ and $C(x_2, R_2) \geq C(y_2, R_2)$ by Assumption 2, then $C(y_1, R_1) \geq C(y_2, R_2)$ which contradicts the assumption of this case. Hence this case cannot happen. Case 3: $C(x_1, R_1) \leq C(x_2, R_2)$ and $C(y_2, R_2) \leq C(y_1, R_1)$. Provided $C(x_2, R_2) \leq C(y_1, R_1)$, then again, the social ordering function corresponds the Rawlsian difference principle. However, if $C(x_2, R_2) \geq C(y_1, R_1)$, then our social ordering function makes the opposite choice. Case 4: $C(x_1, R_1) \leq C(x_2, R_2)$ and $C(y_1, R_1) \leq C(y_2, R_2)$. The opposite choice is always made. Thus, except the Case 4 and the subcase $C(x_2, R_2) \geq C(y_1, R_1)$ in Case 3, our social ordering function corresponds to the Rawlsian difference principle.

1.6 Conclusion

We work in a higher-dimensional economic environment where an indifference hypersurface is the level set of a utility function representing its preference relation. We introduce the concept of curvature that describes the shape of an indifference hypersurface. The IIA assumption is weakened by adding information about the indifference hypersurfaces via curvature. We show that using such information it is possible to construct a rational, anonymous social ordering function that satisfies this weaker IIA condition. Our curvature-based social ordering function coincides with the Rawlsian difference principle under certain circumstances.

Appendix

Gauss-Kronecker curvature of indifference hypersurface as level set of $\ell-$**variable Cobb-Douglas utility function**:

Consider the Cobb-Douglas function $u(x_1, \ldots, x_\ell) = Ax_1^{a_1} \cdots x_\ell^{a_\ell}$ defined for $x_1 > 0, \ldots, x_\ell > 0$, with $A, a_1, \ldots, a_\ell$ positive, $a_1 + ... + a_\ell \leq 1$. Let $\mathbf{N}(w) = \nabla u(w)$ be $(a_1 A x_1^{a_1-1} x_2^{a_2} \ldots x_\ell^{a_\ell}, \ldots, a_i A x_1^{a_1} x_2^{a_2} \ldots x_i^{a_i-1} \ldots x_\ell^{a_\ell}, \ldots, a_\ell A x_1^{a_1} x_2^{a_2} \ldots x_\ell^{a_\ell-1})$ for $w = (x_1, x_2, \ldots, x_\ell) \in I(w, R)$.

A basis of tangent space at w in $I(w, R)$ is

$$v_1 = (a_2 A x_1^{a_1} x_2^{a_2-1} \ldots x_i^{a_i} \ldots x_\ell^{a_\ell}, -a_1 A x_1^{a_1-1} x_2^{a_2} \ldots x_\ell^{a_\ell}, 0, \ldots, 0).$$

$$\ldots$$

$$v_{\ell-1} = (a_\ell A x_1^{a_1} x_2^{a_2} \ldots x_\ell^{a_\ell-1}, 0, 0, \ldots, -a_1 A x_1^{a_1-1} x_2^{a_2} \ldots x_\ell^{a_\ell}).$$

The gradient of $\mathbf{N}$ with respect to each basis vector is

$$\nabla_{v_1}\mathbf{N} = (-a_1 a_2 A^2 x_1^{2a_1-2} x_2^{2a_2-1} x_3^{2a_3} \ldots x_\ell^{2a_\ell}, a_1 a_2 A^2 x_1^{2a_1-1} x_2^{2a_2-2} \ldots x_\ell^{2a_\ell}, 0, \ldots, 0).$$

$$\ldots$$

$$\nabla_{v_{\ell-1}}\mathbf{N} = (-a_1 a_\ell A^2 x_1^{2a_1-2} x_2^{2a} x_3^{2a_3} \ldots x_\ell^{2a_\ell-1}, 0, 0, \ldots, a_1 a_\ell A^2 x_1^{2a_1-1} x_2^{2a_2} \ldots x_\ell^{2a_\ell-2}).$$

Then the numerator and denominator are

$$\left|\det\begin{pmatrix}\nabla_{v_1}\mathbf{N}\\ \vdots\\ \nabla_{v_{\ell-1}}\mathbf{N}\\ \mathbf{N}(w)\end{pmatrix}\right| = A^{2\ell-1}a_1^{\ell-1}a_2\dots a_\ell(a_1+a_2+\cdots+a_\ell)x_1^{(2\ell-1)a_1-\ell}x_2^{(2\ell-1)a_2-2}\dots x_\ell^{(2\ell-1)a_\ell-2}.$$

$$\left|\det\begin{pmatrix}v_1\\ \vdots\\ v_{\ell-1}\\ \mathbf{N}(w)\end{pmatrix}\right| = A^\ell a_1^{\ell-2}x_1^{\ell a_1-\ell}x_2^{\ell a_2-2}\dots x_\ell^{\ell a_\ell-2}(a_1^2x_2^2\dots x_\ell^2+\cdots+a_\ell^2x_1^2\dots x_{\ell-1}^2).$$

$$||\mathbf{N}(w)|| = Ax_1^{a_1-1}x_2^{a_2-1}\dots x_\ell^{a_\ell-1}\left(a_1^2x_2^2\dots x_\ell^2+\cdots+a_\ell^2x_1^2\dots x_{\ell-1}^2\right)^{\frac{1}{2}}.$$

So the Gauss-Kronecker curvature is

$$c(S,w) = A^{\ell-2}\left(\prod_{i=1}^{\ell}a_i\right)\left(\sum_{i=1}^{\ell}a_i\right)\frac{\left(\prod_{i=1}^{\ell}x_i^{(\ell-2)a_i+1}\right)}{\left(\sum_{i=1}^{\ell}a_i^2\prod_{j\neq i}^{\ell}x_j^2\right)^{3/2}}.$$

Chapter 2

Investing in Health: A Macroeconomic Exploration of Short-Run and Long-Run Trade-Offs

2.1 Introduction

Investing in health has been promoted worldwide for economic development. Reports of the World Bank (1993) and the World Health Organization (2001) both advise member countries to invest in health for economic growth and list detailed policy actions. World Bank (1993) (Abstract, p.17, p.51) states that "because good health increases the economic productivity of individuals and the economic growth rate of countries, investing in health is one means of accelerating development." World Health Organization (2001) (p.1–3) reports that "the linkages of health to poverty reduction and to long-term economic growth are powerful, much stronger than is generally understood...the improvements in health would translate into higher incomes and higher economic growth." But does investment in health necessarily cause economic growth and make people better off? And for all countries? Not necessarily. Investing in health involves trade-offs such as short-term loss in consumption and the associated loss in welfare versus longer-term enhanced health, which directly improves the welfare and can potentially generate productivity gains that increase long-run consumption and welfare. This chapter aims to answer these primary questions by disentangling the competing effects.

The most direct effect of investing in health is positive if it makes people healthier, and hence better off. But investing in health requires that resources be reallocated from productive uses outside the health sector, such as the manufacturing sector. Other things equal, in the short run this can reduce the productive capacity of the economy and associated consumption, which harms welfare. In the longer run, however, things are not equal if healthier workers are more productive workers. This indirect effect has two implications. First, it tempers and can even reverse the reduction in manufacturing output. Second, it increases the incentive to invest in capital accumulation because more productive workers may raise the marginal product of physical capital. If so, households react by saving more, and the standard short-term-pain-but-long-term-gain result emerges. Increased saving means less consumption in the short run but more consumption in the long run. As a consequence, if the productivity-enhancing effect of health investment were big enough, there would be a win-win outcome — both more health and more manufactured goods (hence higher economic growth) in the long run.

However, if the productivity-enhancing effect of higher investments in health were "small", then this win-win outcome would not occur. The resulting reduction in total output leads to a reduction in the marginal product of physical capital and less household saving and more consumption. Less short-term pain is incurred, but no long-term gain arises. In this setting of endogenous growth, the productivity growth *rate* is permanently lowered. In this case, the analysis does not support investment in health in terms of economic growth, though still possible to support it if it leads to sufficient welfare gains. The net welfare effect depends on the relative size of the direct health effect and the loss in welfare from less consumption. Therefore, it is critical to investigate these dynamics of investments in health.

The available theoretical literature that attempts to unravel such trade-offs is incomplete, and macro-theoretical studies in health economics are relatively few. The seminal work, Grossman (1972), recognizes the role of health as an investment, and is the first to formally model health as an investment good. In his model of an individual's optimal demand for health, health status is treated as an accumulating stock called “health capital”. However, what Grossman develops is a partial equilibrium of optimal investment on the part of one individual, and he does not consider general equilibrium effects such as the productivity-enhancing effect of health on manufacturing workers. Thus, he is unable to examine the trade-offs of health investment between health and non-health sectors from a broader point of view of an entire economy.

Later literature extends this “health capital” notion in the setting of general equilibrium. However, some studies analyze only the equilibrium effects caused by investment in health and do not consider the transitional dynamics between equilibria. For example, among literature that most focuses on investment in health, Van Zon and Muysken (2001) formally recognize that health sector is a labor-intensive sector and analyze the trade-off of health investment in the form of labor allocation between the manufacturing and the health sector. But they do not find an analytical solution. Van Zon and Muysken (2005) later do find an analytical solution in an extended model, and assess these competing effects at the equilibrium only (i.e., they do not study transitional dynamics). Their numerical results include that an investment in care services of the health labor has a positive welfare effect, and a longer life expectancy unambiguously affects the economy in a negative way. However, these welfare and growth effects are the steady-state effects, the transitional dynamics (both analytical and numerical) between equilibria are still missing. Gong *et al.* (2012) examine an AK growth model and derive analytical equilibrium results of health investment which, as expected, are indefinite due to competing effects illustrated above. To ascertain the definite growth effect of investments in health, they conduct empirical instead of numerical estimations. They run regressions of per capita real GDP growth rate on three types of health investments — the level and growth rate of per capita medical beds, and the ratio of per capita medical beds to per capita real physical capital. They use approximately 20-year Chinese provincial data, which is a rather short panel hence, as they have noted, can only infer short-run growth effect of the health investment. They find that both the level and the growth rate of medical beds have significantly positive influences on the economic growth rate, and that the ratio of medical beds to physical capital has significantly negative effect on the economic growth rate. Thus, both theoretical and empirical findings in Gong *et al.* (2012) about the growth effect of health investment remain ambiguous.

There is no agreement on the weight of health relative to consumption in households utility. For example, Hall and Jones (2007) construct an age-specific overlapping generations model where households' choice is to increase spending on health, which extends life and enhances utility. They assume that the marginal utility of life extension does not decline. As the health spending share of GDP grows along with income, they project that the optimal share in the United States is likely to exceed 30 percent by 2050. The reciprocal of health status enters the utility function as the rate of time preference in their basic model, which allows a cross partial interaction between health and consumption in utility. Whereas in the full dynamic version of their model, Hall and Jones (2007) specify that health status enters the utility function as a separate variable. This requires that they assign a utility weight for health relative to consumption, which they do by drawing on an observation of Nordhaus (2002) regarding the historical value of increases in income and life expectancy in the US. But there is still considerable uncertainty about how households value health relative to consumption. Hence, it is important to assess how sensitive conclusions are to the assumed weights.

This chapter contributes in several ways to the theoretical analysis of health investment in a general equilibrium framework. It constructs an endogenous AK growth model of a two-sector economy. It specifies health and non-health production functions and includes health status in households' utility function (assuming all individual firms are identical, and all households are identical). It explores, both analytically and numerically, the equilibrium shift and transitional dynamics after a once-for-all policy of investment in health (modeled as a reallocation of labor from the non-health to the health sector). It examines the impacts of this policy on households' health status, material consumption, capital accumulation, and discounted utility. The numerically calibrated version of the model indicates for how long the pain of consumption loss lasts, and determines whether these losses are compensated by the long-term gains in health, welfare, and perhaps productivity.

Like the existing literature, the model formally captures relevant trade-offs associated with health investments. Unlike most of the literature which assume private financing of health care (e.g., Gong *et al.* (2012); Halliday *et al.* (2014)), it assumes public financing of the health investment, which accords with the institutional reality that publicly financed health care predominates internationally. We model the health investment as a government policy that reallocates the nation's workforce out of manufacturing and into the health sector. Second, the model provides a sensitivity test for specifications of households utility formulation. On the one hand, our model does not rely on a strong interaction between health and consumption as in Hall and Jones (2007)' basic model. On the other hand, our model assumes Ricardian equivalence (a set-up in which the seniors right at the time of their death are completely replaced by their offspring) as a special case for age-specific overlapping generations as in Hall and Jones (2007)' full model. Third, while our model is in certain respects simpler than the existing literature, it embodies the modern macroeconomic requirement of internal consistency and intertemporal optimization on the part of economic agents, and yet captures the essential trade-offs in a way that is more accessible to policy makers than is the existing theoretical literature. It is hoped that the relative simplicity of our model contributes to the policy debate in countries with a publically-financed health sector such as Canada, the United Kingdom, and the Scandinavian countries. Fourth, we conduct alternative calibrations using different parameter values in the entire range of the productivity-enhancing effect of health. The small value is relevant for developed countries since the aged population who need health services most may well have retired from the workforce, whereas the large value is relevant for developing countries. Thus, the results allow us to address the challenges faced by both developed and developing countries with publicly-financed health sectors.

The remainder of this chapter is divided into several sections. Section 2 presents a two-sector model of health capital in an endogenous AK growth model. Section 3 introduces to the economy a policy shock that diverts labor from the manufacturing to the health sector. It then analyzes the resulting shift from the pre-policy to the post-policy equilibrium. Section 4 further derives the formulae of the transitional dynamics between these two equilibria. Section 5 calibrates the pre-policy equilibrium by choosing values of primitive parameters and pre-determined variables that are consistent with both the model and the economic reality. Section 6 reports the simulation results and sensitivity tests of health parameters. Section 7 concludes and discusses limitations, contributions, and policy implications.

2.2 A Two-Sector Model of Health Capital in an Endogenous AK Growth Model

The model economy involves two sectors: the health sector and the manufacturing sector. In the health sector, we make the following assumptions. We assume that the health production function has only one input, health labor (L), which is measured as a fraction of the total population (normalized to one); that is,

$$\dot{H} = \phi L^{\gamma} - \delta H. \tag{2.1}$$

$\dot{H}$ refers to the change in health status with respect to time. This production function involves positive but diminishing returns: $\phi > 0$ and $0 < \gamma < 1$. For simplicity, the depreciation rate of health status (δ) is assumed to be the same as that of physical capital ($0 < \delta < 1$) as is commonly done in the literature (e.g., Ried (1996); Halliday *et al.* (2014)).

In the non-health manufacturing sector, we assume individual firms are identical. In traditional growth theory, individual firms' Cobb-Douglas production is assumed to have diminishing return to physical capital in manufacturing production. But in the new growth theory, productivity is endogenously enhanced by investment in human capital in the form of knowledge or, as in this paper, health. Such a positive externality requires no private cost of investment to individual firms but social cost from governmental investment in the health sector. Thus, the diminishing-return assumption holds for the individual firms' production but not for the social production. To resolve such a conflict, the growth literature maintains Cobb-Douglas production function with the diminishing return for individual firms (equation (2.2)), while assuming that the productivity of each worker is proportional to the level of human capital, which — in turn — is proportional (θ) to the *aggregate* physical capital stock. For the whole economy, then, there is a linear production function with a constant return for the society (equation (2.4)). In particular, the individual (denoted with the "i" subscript) firms' production function has the Cobb-Douglas form in which the effective labor input ($q(1-L)$). The extension to the standard Romer (1990)'s AK growth model that is involved here is that worker productivity is affected by both education and health ($\beta > 0$); that is,

$$Y_i = K_i^{\alpha}\big(q\,(1-L_i)\big)^{1-\alpha}, \tag{2.2}$$

$$q = \theta K H^{\beta}. \tag{2.3}$$

We assume that health raises labor productivity but with diminishing returns ($0 < \beta < 1$). The scale parameter θ has no restriction on its value. Equations (2.2) and (2.3) can be combined to yield, at the aggregate level, the society-level production function in the manufacturing sector

$$Y = AK, \tag{2.4}$$

where

$$A = \theta^{1-\alpha} H^{\beta(1-\alpha)}(1-L)^{1-\alpha}. \tag{2.5}$$

Firms in the manufacturing sector pay the interest rate (r) to rent physical capital and the wage rate (w) to rent labor. Input levels are chosen to maximize profits $\pi = Y - (r+\delta)K - w(1-L)$ (total output (sales) less costs paid as the capital income, depreciation, and the labor income). Thus, they hire each factor up to the point that the marginal product equals the rental price; that is,

$$\frac{\alpha Y}{K} = \alpha A = r + \delta, \tag{2.6}$$

which is an individual firm's optimal hiring rule for physical capital; and

$$\frac{(1-\alpha)Y}{1-L} = w, \tag{2.7}$$

which is an individual firm's optimal hiring rule for labor.

Next we define the economy's resource constraint

$$Y = C + \dot{K} + \delta K + G = w(1-L) + rK + \delta K, \tag{2.8}$$

where G is government spending on non-health programs (not transfers). As economic growth proceeds, we assume that the government increases G so that its ratio to total manufacturing sector output, $z = G/Y$, remains constant. The wage rate (w) for health and non-health labor is assumed to be the same. The interest rate (r) is the rate of return on capital, net of depreciation. These relationships state that the total amount of labor and capital income equals the sum of total spending by the private households' consumption and investment, plus the government's spending on non-health programs.

We now turn to the final identity in the model—the government's budget constraint. We note that, in reality, the optimal choice of investing in health is a mixed decision by both individual households and the government. But countries with dominant publically-financed health sector such as Canada and the United Kingdom can control the health investment through both financing policies and by controlling the admissions of domestic medical schools and immigrations of international medical graduates. Moreover, the health sector is a labor-intensive service sector. Thus, we assume that in this model the choice of optimal health investment is made solely by the government. The government's budget constraint is

$$G + wL = \tau\big(w + rK\big), \tag{2.9}$$

which states that the government collects taxes imposed on both labor and capital income (net of depreciation expenses) at the same tax rate (τ), and uses this revenue to pay the wages of the health labor and to pay for non-health governmental programs (G). Note that identity (2.9) can be rewritten using the optimal hiring rules as

$$z + \frac{(1-\alpha)L}{1-L} = \big(\frac{\tau}{1-\tau}\big)(1 - \frac{\delta}{A} - z). \tag{2.10}$$

Households maximize utility, which involves additive terms of consumption and health status in the logarithm form (a simple and standard way of imposing diminishing marginal utility), and a constant rate of time preference, ρ; that is,

$$U = \int_0^\infty e^{-\rho t}(lnC_t + \xi \ln H_t)dt, \tag{2.11}$$

subject to

$$C + \dot{K} = (1-\tau)(rK + w). \tag{2.12}$$

This optimization yields the familiar Ramsey (1928) consumption function

$$\frac{\dot{C}}{C} = r(1-\tau) - \rho. \tag{2.13}$$

We use household utility as our measure of welfare when performing our normative analysis later in this chapter.

We now have established the model of an economy with a health sector and a manufacturing sector, which specifies how labor enters into each sector and indicates a trade-off between health labor and manufacturing labor. Next, we need to write the model as a dynamic system that compactly represents this economy. We then verify whether this system converges to an equilibrium; and if yes, analyze the qualitative and quantitative properties of the equilibrium before any policy shock is implemented.

2.3 Existence of Pre-Policy Equilibrium and Shift to Post-Policy Equilibrium

In this section, we first explain the pre-policy equilibrium. Then we introduce an exogenous health labor policy that reallocates some of the nation's workforce from the manufacturing sector to the health sector, doubling health labor in a once-for-all fashion. We analyze the shift in the economy's full equilibrium, and later the adjustment path between full equilibria.

2.3.1 Pre-Policy Equilibrium

To verify the existence of the pre-policy equilibrium, first we define the ratio of consumption to physical capital, $x = C/K$. We are interested in both consumption and the physical capital stock independently. But since both of them rise to infinity as time proceeds, only the ratio of the two has a fixed value to which the dynamic system of the economy can converge.

We summarize the model as a system of four equations. First, we have

$$\frac{\dot{x}}{x} = \big[\alpha(1-\tau) - (1-z)\big]A - \rho + x + \tau\delta, \tag{2.14}$$

which is obtained by substituting equation (2.13) and the left-hand side of equation (2.8) into the time derivative of x. The other three equations are restatements of equations already presented

$$\frac{\dot{H}}{H} = \phi L^{\gamma} H^{-1} - \delta, \tag{1}$$

$$A = \big[\theta(1-L)\big]^{1-\alpha} H^{\beta(1-\alpha)}, \tag{5}$$

$$\tau = \frac{z + (\frac{1-\alpha}{1-L})L}{1 - \frac{\delta}{A} + (\frac{1-\alpha}{1-L})L}. \tag{10}$$

These four equations, (2.14, 1, 5, and 10), involve five endogenous variables — the time rate of change in H and x, and the current values of x, A, and τ. Because the two ordinary differential equations (1) and (2.14) involving H and x generate the two-dimensional dynamic system that represents this economy, we focus on these two equations to find the equilibrium of the economy. At each point in time, when the values of A, τ, and H are given, these two equations determine x, $\dot{x}$, and $\dot{H}$, since x is a jump variable, while H is not. The reason for this difference is that x is determined by the households' *forward-looking* consumption plan; whereas the change in health status follows from a historically based (*backward-looking*) accumulation identity.

To close the model, we need a fifth restriction together with the four equations above to solve for the five endogenous variables. The standard procedure it to assume whatever it takes to ensure unique convergence to a full equilibrium (that is, to apply Samuelson (1941)s Correspondence Principle). In this case, we need to check whether the model involves a saddle path; and if so, to assume that the initial value of x is determined by requiring the economy be on that saddle path.

Now we verify whether such a saddle equilibrium exists, then specify the saddle-path equation, if it exists. We examine the equilibrium and its neighborhood by linearization, as widely done in the literature (Scarth, 2014). Taking a linear approximation (the total differential) of equations (1) and (2.14), we get

$$d\dot{H} = [-\delta]dH + [\phi\gamma L^{\gamma-1}]dL, \tag{2.15}$$

$$\begin{aligned} d\dot{x} = &\left[\frac{(\alpha(1-\tau)-(1-z)+rq_2)\beta(1-\alpha)Ax}{H}\right]dH + [x]dx \\ &+ \left[\left(\big(\alpha(1-\tau)-(1-z)+rq_2\big)\big(\frac{\alpha-1}{1-L}\big)A - rq_1\right)x\right]dL, \end{aligned} \tag{2.16}$$

where

$$q_1 = \frac{\left(\frac{1-\alpha}{1-L}\right)\left(1+\frac{L-\tau}{1-\tau}\right)}{\frac{1-\alpha}{1-L}+\frac{r}{A}},$$

$$q_2 = \frac{\frac{\tau\delta}{A^2}}{\frac{1-\alpha}{1-L}+\frac{r}{A}}.$$

We rewrite the resulting two equations in a matrix form

$$\begin{bmatrix} d\dot{H} \\ d\dot{x} \end{bmatrix} = \begin{bmatrix} -\delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix} \begin{bmatrix} dH \\ dx \end{bmatrix} + \begin{bmatrix} \phi\gamma L^{\gamma-1} \\ \left(\left(\alpha(1-\tau)-(1-z)+rq_2\right)\left(\frac{\alpha-1}{1-L}\right)A - rq_1\right)x \end{bmatrix} dL. \tag{2.17}$$

When there is no policy shock, that is, $dL = 0$, the dynamic system (2.17) becomes

$$\begin{bmatrix} d\dot{H} \\ d\dot{x} \end{bmatrix} = \begin{bmatrix} -\delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix} \begin{bmatrix} dH \\ dx \end{bmatrix}.$$

Define the matrix of this pre-policy system as

$$B = \begin{bmatrix} -\delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix}. \tag{2.18}$$

The determinant of matrix B is the product of the two eigenvalues, $-\delta x < 0$, where one eigenvalue is positive and the other negative. This verifies that there exists a saddle equilibrium for the system. As noted, this is necessary for a unique convergence since there is one sticky variable, H, and one jump variables, x. Moreover, because the negative real eigenvalue (here it is $-\delta$) is the adjustment speed of the dynamic system (Smale *et al.*, 2012), we define $v = -\delta$ as the velocity of our system adjusting along the saddle path. Given the linear approximation, all endogenous variables move with this adjustment speed between the initial and final equilibria.

2.3.2 Exogenous Investment in Health Sector

Having verified the existence of a saddle equilibrium and found the adjustment speed along that saddle path, we now introduce the exogenous health labor policy that reallocates the manufacturing labor to the health sector. We assume for simplicity that the health labor policy is a once-for-all event that households and firms do not anticipate. Given our equilibrium is a saddle, x jumps to put the economy on the new saddle path after the policy shock. Figure (2.1) depicts the trajectory of the economy following the policy. A sudden jump occurs from the pre-policy equilibrium (point 1) to point 2 on the new saddle path. Then, as time passes following this initial change, the economy tracks along the new saddle path from point 2 to 3. Point 3 is the post-policy equilibrium. As noted in the introduction, this short-term pain (lower consumption initially) followed by the long-term gain (higher consumption in the post-policy full equilibrium) — as shown by the curved arrow label on the vertical axis — is just one possibility. To establish what outcome is likely, we must examine a calibrated version of the model.

Given parameter values and initial values of τ, L and H, we can derive the effects of the health labor policy on the equilibrium values of health status (dH^*/dL) and the ratio of consumption to physical capital (dx^*/dL). To do so, we set $\dot{H}$ and $\dot{x}$ to zero (i.e., $\dot{H} = 0$, $\dot{x} = 0$). Since $d\dot{H}$ and $d\dot{x}$ are defined as $(\dot{H} - \dot{H}^*)$ and $(\dot{x} - \dot{x}^*)$, and since $\dot{H}^* = \dot{x}^* = 0$, we set $d\dot{H}$ and $d\dot{x}$ to zero in equation (2.17) to get

$$\begin{bmatrix} \delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix} \begin{bmatrix} dH^* \\ dx^* \end{bmatrix} = \begin{bmatrix} \phi\gamma L^{\gamma-1} \\ \left(\left(\alpha(1-\tau)-(1-z)+rq_2\right)\left(\frac{1-\alpha}{1-L}\right)A + rq_1\right)x \end{bmatrix} dL. \tag{2.19}$$

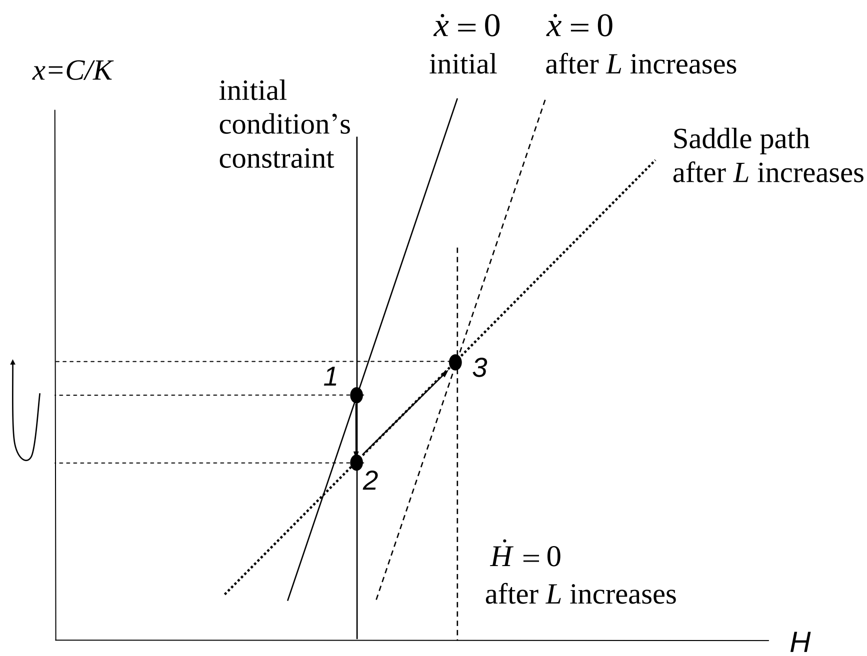


Figure 2.1: Phase diagram shows both saddle-path equilibrium and one possible outcome: short-term pain and long-term gain following an unexpected but permanent expansion of the health sector.

Denote the left-hand-side matrix as

$$B' = \begin{bmatrix} \delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix}, \tag{2.20}$$

based on which, we can follow Cramer's rule. To do so, we define two matrices that are associated with B'; that is,

$$B_x = \begin{bmatrix} \delta & \phi\gamma L^{\gamma-1} \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & \left(\left(\alpha(1-\tau)-(1-z)+rq_2\right)\left(\frac{1-\alpha}{1-L}\right)A+rq_1\right)x \end{bmatrix}, \tag{2.21}$$

$$B_H = \begin{bmatrix} \phi\gamma L^{\gamma-1} & 0 \\ \left(\left(\alpha(1-\tau)-(1-z)+rq_2\right)\left(\frac{1-\alpha}{1-L}\right)A+rq_1\right)x & x \end{bmatrix}. \tag{2.22}$$

By Cramer's rule based on the determinants of these matrices, we obtain the equilibrium shifts dH^* and dx^* following a change in the size of the health sector workforce dL

$$\frac{dH^*}{dL} = \frac{|B_H|}{|B'|}, \tag{2.23}$$

$$\frac{dx^*}{dL} = \frac{|B_x|}{|B'|}. \tag{2.24}$$

Further by assumptions in section 2, $\phi > 0$, $0 < \delta < 1$, $0 < \gamma < 1$, and $0 < L < 1$, we derive that $dH^*/dL = \phi\gamma L^{\gamma-1}/\delta > 0$. This means that an increase in the health labor raises the equilibrium value of health status. However, dx^*/dL has a very complicated expression, and its sign is indeterminate, hence a need for calibration in Section 5.

Given $dH^* = H^* - H_0$ and $dH^* = (|B_H|/|B'|)\ dL$, the post-policy equilibrium value of health status is

$$H^* = H_0 + dH^* = H_0 + \frac{|B_H|}{|B'|}dL, \tag{2.25}$$

where H_0 is the initial value of the health status before policy (the pre-policy equilibrium value).

Similarly, given $dx^* = x^* - x_0$ and $dx^* = (|B_x|/|B'|)\, dL$, the post-policy equilibrium value of x is

$$x^* = x_0 + dx^* = x_0 + \frac{|B_x|}{|B'|} dL, \tag{2.26}$$

where x_0 is the initial value of x before policy (the pre-policy equilibrium value).

2.4 Transitional Dynamics Between Pre-Policy and Post-Policy Equilibria

We have found the pre-policy and the post-policy equilibria. However, knowing only the equilibria before and after the policy is insufficient. Recall our goal is to compare the short- and long-term pain and gain throughout the time horizon following the implementation of the health investment policy. We now derive the transitional dynamics from the pre-policy to the post-policy full equilibria at each point in time for variables of interest: health status (H_t), the ratio of consumption to physical capital (x_t), the physical capital (K_t), consumption (C_t), and discounted summed utility (U_t).

2.4.1 Calculate H_t

Figure 2.2 illustrates the pattern of H_t after the policy shock. Given our assumption that the system is linearized, we analyze the transitional dynamics in the neighborhood of the post-policy full equilibrium. Solving the first-order differential equations (1) and (2.14) for H_t and x_t, and setting any further changes in L to zero, we have

$$d\dot{H} = vdH \text{ or } \dot{H} = -\delta(H - H^*) \tag{2.27}$$

which implies

$$H_t = H^* - (H^* - H_0)e^{vt}. \tag{2.28}$$

By substituting H^* from equation (2.25), H_0, and v into equation (2.28), we obtain H_t for all time t, which is the transitional dynamics of health status between the two full equilibria.

2.4.2 Calculate x_t

Similarly, because both variables H and x in the dynamic system share the same speed of adjustment along the saddle path, we have the solution for x_t

$$x_t = x^* - (x^* - x_j)e^{vt}, \tag{2.29}$$

where x_j is the initial value of x after the labor policy has been implemented (the post-policy initial value) since x is a jump variable. Notice that $x_j < x_0$ and $x_j > x_0$ are both possible. In other words x_t can jump either down or up after the policy shock, as illustrated in Figure 2.3. Also, the new equilibrium value x^* can be higher

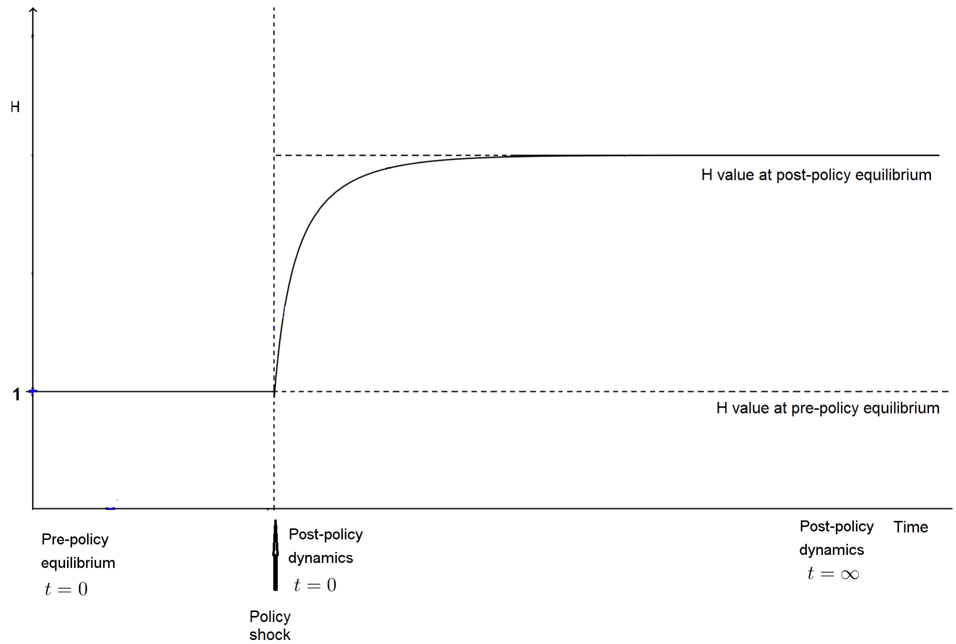


Figure 2.2: Illustrated transitional dynamics of health status before and after a health labor policy shock.

or lower than x_0. Such ambiguities require numerical simulations in Section 5.

By substituting equation (2.26), x_j, and v into equation (2.29), we can get x_t if we know x_j. To find x_j, we rely on the fact that for a linearized system the equation of saddle path must be linear; that is,

$$x^* - x_t = b(H^* - H_t), \tag{2.30}$$

By substituting equation (2.26), x_j, and v into equation (2.29), we can get x_t if we know x_j. To find x_j, we rely on the fact that for a linearized system the equation of saddle path must be linear; that is,

$$x_j = x^* - b(H^* - H_0). \tag{2.31}$$

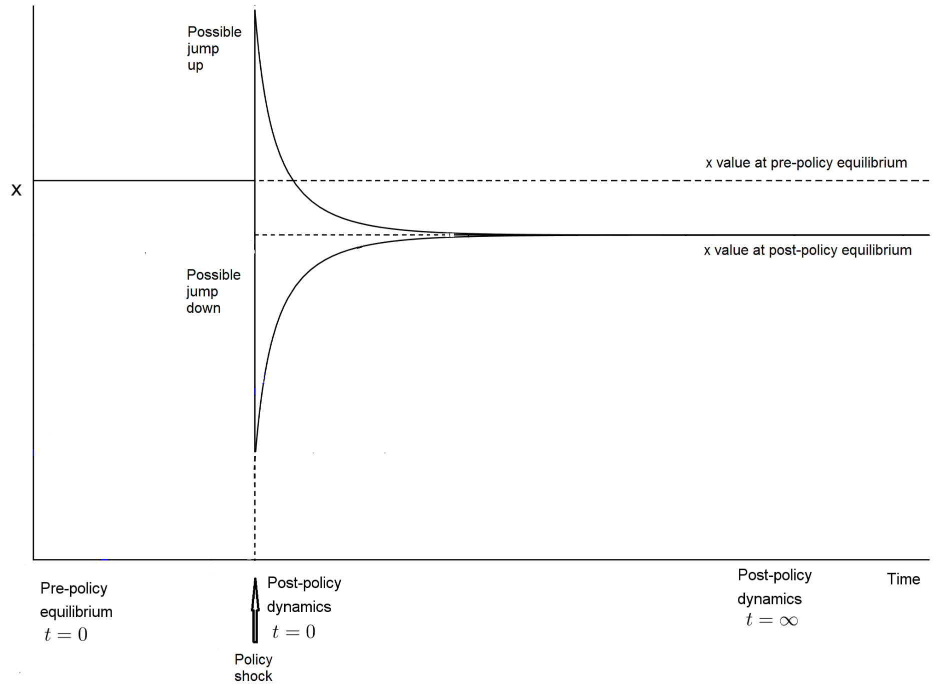


Figure 2.3: Illustrated transitional dynamics of the ratio of consumption to physical capital before and after a health labor policy shock where its initial value after policy can jump either down or up.

To find b, denote $l := H^* - H_0$, we rewrite equations (2.28) and (2.30):

$$H^* - H_t = (H^* - H_0)e^{vt} = le^{vt}, \tag{2.32}$$

$$x^* - x_t = (x^* - x_0)e^{vt} = ble^{vt}. \tag{2.33}$$

We can rewrite our system (2.17) without the exogenous variable dL (which is appropriate for all time periods beyond $t = 0$, since L changes only at $t = 0$) as

$$\begin{bmatrix} \dot{H} \\ \dot{x} \end{bmatrix} = \begin{bmatrix} -\delta & 0 \\ \frac{\left(\alpha(1-\tau)-(1-z)+rq_2\right)\beta(1-\alpha)Ax}{H} & x \end{bmatrix} \begin{bmatrix} dH \\ dx \end{bmatrix}. \tag{2.34}$$

By substituting both equations (2.32) and (2.33) and their time derivatives into equation (2.34), we find that

$$b = \frac{\big(\alpha(1-\tau) - (1-z) + rq_2\big)\beta(1-\alpha)A\bar{x}}{(-\delta - x)H}. \tag{2.35}$$

Finally, by substituting b back to equation (2.31), we get x_j. Then, by substituting x_j, x^* and v back to equation (2.29), we obtain x_t for all time t.

2.4.3 Calculate K_t

Given that we know $x_t = C_t/K_t$ for all time t, to find C_t we first need to find K_t. This is straightforward since, at each point in time, both A_t and x_t are given. We know that

$$\frac{\dot{K}}{K} = (1-z)A - x - \delta,$$

which implies

$$K_t = e^{((1-z)A-x-\delta)t} K_0, \tag{2.36}$$

where K_0 is the initial capital stock (the pre-policy equilibrium value). Equation (2.36) can be used, along with the values we know for x_t and A_t (from equations (2.29) and (5)) to calculate K_t for all t beyond $t = 0$.

2.4.4 Calculate C_t

Finally, to get C_t, we substitute x_t from equation (2.29) and K_t from equation (2.36) into

$$C_t = x_t K_t. \tag{2.37}$$

Corresponding to x_j, C_j is the initial value of jump variable C after the labor reallocation policy (the post-policy initial value); that is,

$$C_j = x_j K_0. \tag{2.38}$$

2.4.5 Calculate U_t

After substituting C_t and H_t into the discrete-time version of equation (2.11):

$$U_t = \sum_{t=0}^{f} \left(\frac{1}{1+\rho}\right)^t \Big[\ln C_t + \xi \ln H_t\Big], \tag{2.39}$$

we obtain U_t for any chosen time horizon of f years into the future.

2.5 Calibration of Pre-Policy Equilibrium and Sensitivity Tests of Health Parameters

To generate numerical simulations, we need to calibrate the model—in particular, the pre-policy equilibrium—and then examine the equilibrium shift caused by the health labor policy and conduct sensitivity tests. We choose plausible values of primitive parameters and pre-determined variables by two criteria. First, they must satisfy all equations in section 2 when the economy is in its original full equilibrium (recall $\dot{H} = 0$, $\dot{x} = 0$, and $\dot{K}/K = \dot{C}/C = g$). Second, broadly speaking, they should match real world observations. Below we first show how we derive all but three of the primitive parameters and pre-determined variables at the pre-policy equilibrium, denoted with a "0" subscript. These values chosen are consistent with a large amount of applied macroeconomic literature. Then we derive health-related parameters that are perhaps more controversial because there is less evidence available in the macro-oriented health economic literature.

First, we calibrate parameters and pre-determined variables associated with production. We assume that depreciation rates (δ) of physical capital and health are identical at 4% per year, so δ is 4%. This is somewhat arbitrary but still consistent with existing macroeconomic literature (e.g., Halliday *et al.* (2014)). The total output of the economy (Y_0) is normalized to 2. Historically, the annual real per capita economic growth rate (g_0) of countries such as Canada, the United Kingdom, and Australia has been around 2% (OECD Statistics, 2014). Hence, that is what we assume. We further assume that investment in physical capital as a fraction of the total manufacturing output (I_0/Y_0) equals 20%, which matches the historical pattern of domestic capital investment as a percentage of GDP for countries such as Canada, the United Kingdom, and the United States (World Bank (2015) on average over the business cycle). To derive K_0, first we divide the investment identity by Y_0; that is,

$$\frac{I_0}{Y_0} = \frac{\dot{K}}{Y_0}\frac{K_0}{K_0} + \delta\frac{K_0}{Y_0},$$

then substitute $Y_0 = 2$ and $\dot{K}/K_0 = g_0$, we get $K_0 = I_0/(g_0 + \delta)$. By substituting values of δ, g_0, and I_0, we obtain K_0 as a value of 6.667, implying a realistic capital/output ratio of 3.333, and $Y_0/K_0 = A_0$ as 0.3. The capital share of national income (α) is conventionally assumed to be 1/3, and we follow that convention here.

We normalize the size of the entire workforce to 1. For Canada, health labor accounts for 5% of the total workforce (Statistics Canada, 2014), so we let the initial

L value (L_0) be 0.05. The health investment policy is to divert labor from the manufacturing to double the health sector's share from 0.05 to 0.10, that is, $dL = 0.05$. By substituting values of δ, Y_0, A_0, α, and L_0 into firms' optimal hiring equations (2.6 and 2.7), we obtain the initial values for annual interest rate (r_0) as 0.06 which is quite reasonable and wage rate (w_0) as 1.404 which is required to have labor receive two-thirds of national income (a realistic proportion).

We next calibrate parameters and pre-determined variables associated with the government. We assume the income tax rate (τ_0) to be 0.3 which is broadly representative of developed countries such as Canada, the United Kingdom, and many Western European countries in terms of the ratio of tax revenue to GDP (OECD Statistics (2014)). Then we check for the value of non-health governmental spending that is implied by the government budget constraint (equation 2.9). Deducting the wage paid to health labor from the total tax revenue, $\tau_0(r_0K_0 + w_0) - w_0L_0$, gives G_0 as 0.471, and hence its share of the total manufacturing output is 0.235.

Finally, we calibrate parameter values and pre-determined variables associated with households. To derive the time preference rate (ρ), we substitute $(\dot{C}/C_0) = g_0$ into equation (2.13) and get $\rho = r_0(1 - \tau_0) - g_0$ which turns to be 0.022. Although this value is slightly below the value of 0.04 that is often assumed in the applied macroeconomic literature, it is still in the acceptable range, and we must assume this value for model consistency. We normalize health status as an index with a starting value (H_0) of 1. There are many measurements or proxies of health level in the literature (e.g., life expectancy at birth or any other specific age, infant mortality rate, stature). By indexing health level, we can obtain a much more general measurement, which can then be transformed back to the level of any specific health measure. To derive C_0, we substitute values of Y_0, I_0, and G_0 into the economy's resource constraint $Y_0 = C_0 + I_0 + G_0$, and this gives C_0 as 1.129 which accounts for 56.45% of the total manufacturing output. Since households' consumption is in the range of 60% of GDP in most Western economies, we are quite comfortable with this value. Immediately $x_0 = C_0/K_0$ turns to be 16.94%. Table 2.1 summarizes these calibrated primitive parameters and pre-determined variables.

Table 2.1: Calibrate the baseline equilibrium: Primitive parameters and pre-determined variables

δ	depreciation rate	0.04
g_0	annual economic growth rate	0.02
Y_0	total output	2
I_0	ratio of investment to total output	0.4
K_0	physical capital	6.667
A_0	productivity of non-health goods' production	0.3
α	return to physical capital in non-health goods' production	1/3
L_0	fraction of total labor force in health sector	0.05
r_0	real interest rate excluding the depreciation rate	0.06
w_0	wage	1.404
τ_0	income tax rate	0.3
G_0	governmental spending on non-health sector	0.471
ρ	time preference rate	0.022
H_0	index of health level	1
C_0	consumption of non-health goods	1.129
x_0	ratio of consumption to physical capital	16.94%

Notes: Note: Pre-determined variable at the pre-policy equilibrium is denoted with a "0" subscript.

As mentioned above the three health-sector parameters (β, ξ, and γ) are less well determined, because they are less often reported in the literature. Thus, a full explanation of how we calibrated values are chosen is required. Tompa (2002) systematically reviews the empirical evidence about how much the increase in labor productivity is due to workers being healthier (e.g., Knowles and Owen (1997), Bloom *et al.* (2004)). Unfortunately, these estimates vary considerably across studies depending on the specific data, methods, and health measures used. Such a variety can be rationalized by one argument in Barro and Sala-i Martin (1995) (p.432): "it is likely that life expectancy has such a strong, positive relation with growth because it proxies for features other than good health that reflect the desirable performance of a society. For example, higher life expectancy may go along with better work habits and a higher level of skills (for given measured values of per capita product and years of schooling)." Given this considerable uncertainty concerning the productivity-enhancing effect of health (our parameter β), we examine the entire range of theoretical possibility given the assumption of diminishing return $0 < \beta < 1$. We report three values in the range — small as 0.05 (which may better reflect the experience of developed countries since their aged population who need health services most may have retired from the labor force); large as 0.95 (in contrast, perhaps better representing the experience in developing countries); and moderate as 0.50.

We now choose a calibrated value for the relative weight of health in the utility function, ξ. We start with the report of Jones (2013) (p.1) that "during the twentieth century, life expectancy in the United States rose from less than 50 years to 77 years, while average incomes rose by about a factor of 7." Jones notes that Nordhaus (2002) estimates that individuals have evaluated both these favorable outcomes as generating about the same change in their utility. We can exploit these observations as follows.

Consider our instantaneous utility function in equation (2.11) which implies

$$\frac{\Delta U}{U} = \frac{\Delta C}{C} + \xi \frac{\Delta H}{H}.$$

If we set $\Delta U/U = 0$, we have $\Delta C/C = -\xi \Delta H/H$. Given that Nordhaus (2002) and Jones (2013) have considered $\Delta H/H = 1/2$ and $\Delta C/C = 7$, we estimate ξ as 14. Given the considerable uncertainty regarding the true value of this parameter, we conduct a sensitivity test regarding its value. We view the value of 14 as likely being at the high end of plausible values because it is based only on health changes that lengthen life. In reality, many health improvements produce improvements in quality of life while not extending it. As a sensitivity test, therefore, we test an alternative, smaller value for this parameter, $\xi = 5$.

Now we explain how we pick a value for the parameter (γ) that defines how the investment in health increases overall health status using two different methods. One way to recover $\gamma = dH/H$ (which comes from $\gamma dL/L = dH/H$ where both L and dL are 0.05) is via the first-order connection between health labor L and expenditure on health labor as a percentage of GDP, which in our model is $Y + wL$. We follow macroeconomic convention by measuring unsold government-provided goods (in this case, health care) at cost. Thus, health care expenditure expressed as a proportion of GDP, denoted as E, is $wL/(Y + wL)$. Taking the differential, we have

$$dE = \frac{Yw}{(Y + wL)^2} dL,$$

which implies

$$\frac{dE}{E} = \frac{Y}{Y + wL} \frac{dL}{L},$$

and results in $dE/E = 0.588$ after substituting in the assumed parameter values for Y, w, and L from our Table 2.1.

Now, let us assume that the change in health is proportional to the change in health expenditure; that is,

$$\frac{dH}{H} = m\frac{dE}{E},$$

and then use actual data to estimate m. Between 1979 and 2007, we note that following facts about Canada (obtain from OECD (2015)). During this period, total health expenditure rose from 6.791% to 10.037% of GDP (so dE/E is 0.478), and life expectancy at birth rose from 75.1 to 80.7 years (so dH/H is 0.075). Then m is estimated to be $0.075/0.478 = 0.156$. Combining the two relationships that have been our focus in this and the previous paragraphs, we have

$$\begin{aligned}\frac{dH}{H} &= \left(\frac{mY}{Y + wL}\right)\frac{dL}{L} \\ &= (0.156)(0.588)\frac{dL}{L} \\ &= 0.09\frac{dL}{L}.\end{aligned}$$

Finally, since the full-equilibrium version of equation (2.1) implies

$$\frac{dH}{H} = \gamma\frac{dL}{L}, \tag{2.40}$$

this reasoning and evidence justify our taking γ to be 0.09.

Another way to find the γ value follows from a very stylized model in Van Zon and Muysken (2005) and gives γ as 0.03 (see Appendix). We simulate with both values and report both results below.

Table 2.2 summarizes the two sets of key health parameters β, ξ, and γ. We shall refer to them as the "large" and "small" parameter sets. Given the literature from which the β values were taken we interpret the "large" set as more relevant for developing economies and the "small" as more relevant for developed economies. To ensure that all equations are satisfied for both parameter sets, two other scale parameters need to be determined. In particular, ϕ and θ are derived from these three parameters via equations (1) and (5) respectively. Set equation (2.1) zero, substitute values of H_0, L_0, δ, and γ, we obtain the value of coefficient ϕ. By substituting into equation (2.5) the values of A_0, H_0, L_0, α, and β, we get the value of the coefficient θ. We report simulation results in the next section that emerge with both parameter sets.

Table 2.2: Range of health parameter values

	β	ξ	γ	ϕ	θ
Large	0.95	14	0.09	0.052	0.173
Small	0.05	5	0.03	0.044	0.173
Middle	0.50	14 or 5	0.09	0.052	0.173
	0.50	14 or 5	0.03	0.044	0.173

2.6 Results

We first report the pre-policy equilibrium of the economy. We then report the equilibrium shifts and transitional dynamics in the short (one year), medium (ten years), and long (thirty-five years) run caused by the policy that reallocates labor from the manufacturing to the health sector, doubling the health labor from 5% to 10% of the nation's workforce. We report two types of comparisons — Pre-policy vs. Post-policy using large parameter set, and Post-policy using large parameter set vs. Post-policy using small parameter set. We then report how the conclusions are sensitive to the three health parameters — the households' valuation of health status versus material consumption (ξ), the productivity-enhancing effect of health (β), and the effectiveness of additional labor in generating increased health status (γ). Finally, we report a simulation particularly suitable for developed countries like Canada where the productivity-enhancing effect of health is small because the people who enjoy health services in rich economies may, to a large extent, be the elderly who have retired from the workforce.

2.6.1 Pre-Policy Equilibrium

Table 2.3 shows the calibrated pre-policy equilibrium of the economy. Health labor (L_t) remains at 5% of the nation's workforce. Health (H_t) is at the normalized index value 1. The ratio (x_t) of consumption to physical capital remains at 16.94%. The productivity (A_t) is 0.3, and the income tax rate (τ_t) is 30%. The interest rate (r_t) remains at 6%, and economic growth rate (g_t) stays at 2%. While the ratio x_t is fixed, its denominator (K_t) and numerator (C_t) keep growing, starting at 6.667 and 1.129, reaching at 6.801 and 1.152 after one year, and 13.425 and 2.274 after thirty-five years.

2.6.2 Post-Policy Equilibrium and Transitional Dynamics Using Large Parameter Set

After introducing the policy shock that doubles the health labor from 5% to 10% of the labor force, the economy's equilibrium shifts from the pre-policy toward the

Table 2.3: Pre-policy equilibrium

	Pre-policy equilibrium	Pre-policy dynamics					
	$t = 0$	$t = j$	$t = 1$	$t = 10$	$t = 35$	$t = 50$	$t = \infty$
L_t	5%	5%	5%	5%	5%	5%	5%
H_t	1	1	1	1	1	1	1
x_t	16.94%	16.94%	16.94%	16.94%	16.94%	16.94%	16.94%
A_t	0.3	0.3	0.3	0.3	0.3	0.3	0.3
τ_t	30%	30%	30%	30%	30%	30%	30%
r_t	6%	6%	6%	6%	6%	6%	6%
g_t	2%	2%	2%	2%	2%	2%	2%
K_t	6.667	6.667	6.801	8.143	13.425	18.122	∞
C_t	1.129	1.129	1.152	1.379	2.274	3.069	∞

Notes: Note: j denotes jump, which means the initial value if a policy is implemented.

post-policy equilibrium. Table 2.4 reports the transitional dynamics between the two equilibria using the large parameter set. H_t increases gradually, reaching 1.021 after ten years, 1.048 after thirty-five years, and eventually converging to 1.064. If the health measure were life expectancy at birth, this would imply that after doubling the health labor force, the households' life expectancy at birth increases from 75 to 76.6 and 78.6 years after ten and thirty-five years of policy respectively, and finally reaches 79.8 years at the post-policy equilibrium.

x_t suddenly jumps down from 16.94% to 16.67% as the post-policy initial value, then slowly increases to 16.69% after one year, 17.04% after thirty-five years, and converges to 17.17%, higher than what occurs without the policy. Such a sudden drop comes from the households' forward-looking cut in consumption right after the announcement of the policy. And the long-term increase in x_t may result from higher labor productivity (confirmed below) and hence more manufacturing output for more consumption.

Table 2.4: Equilibrium shift and transitional dynamics caused by the policy that reallocates labor from manufacturing to health sector to double health labor from 5% to 10% of the nation's workforce using the **large** parameter set

	Pre-policy equilibrium	Post-policy dynamics					
	$t=0$	$t=j$	$t=1$	$t=10$	$t=35$	$t=50$	$t=\infty$
L_t	5%	10%	10%	10%	10%	10%	10%
H_t	1	1	1.003	1.021	1.048	1.056	1.064
x_t	16.94%	16.67%	16.69%	16.83%	17.04%	17.10%	17.17%
A_t	0.3	0.289	0.290	0.293	0.298	0.299	0.301
τ_t	30%	33.07%	33.07%	33.01%	32.93%	32.91%	32.88%
r_t	6%	5.65%	5.66%	5.78%	5.94%	5.98%	6.03%
g_t	2%	1.58%	1.59%	1.67%	1.78%	1.81%	1.85%
K_t	6.667	6.667	6.765	7.814	12.321	16.381	∞
C_t	1.129	1.111	1.129	1.315	2.100	2.801	∞

Notes: Note: j denotes jump, which means the post-policy initial value.

A_t immediately drops from 0.3 to 0.289 due to the sudden withdrawal of manufacturing labor. It then slightly increases throughout the time and converges to 0.301, negligibly higher than 0.300 without the policy. This suggests that when health parameters (in particular, the productivity-enhancing effect of health, β) is large, the productivity increases though the increase is negligible in the long run.

τ_t immediately jumps up to 33.07% because the government needs to raise more tax revenue to finance the publicly-funded expansion of health labor. The tax rate then slightly decreases and converges to 32.88%, still higher than the pre-policy initial value 30%. This is because the policy is permanent so that the government has to continue raising enough revenue to pay these new health professionals.

r_t immediately decreases from 6% to 5.65%. This outcome follows from the optimal hiring rule for physical capital — capital's rate of return falls because its productivity, A_t, falls. r_t then slowly goes back to 6.03% at the post-policy equilibrium, which corresponds to the slightly enhanced productivity A_t as 0.301 in the long term.

g_t falls considerably from 2% to 1.58% right after the policy, which results from the decrease in interest rate and increase in income tax rate. Both these developments decrease the households' incentive to save. But g_t largely recovers to 1.81% after half of a century and eventually converges to 1.85%, though still lower than the pre-policy initial growth rate 2%. This is because, in the long term, the decrease in the interest rate and the increase in the tax rate are both less than in the short term. Recall that the productivity-enhancing effect of health (β) is 0.95. This result means that even if health brings a nearly constant return in labor productivity, investment in health still slows down economic development; although more than half of the slowdown in the short run is recovered in the long run.

K_t rises from 6.667 to 6.765 after one year, 7.814 after ten years, and 12.321 after thirty-five years of introducing the policy. Comparing to its counterparts without the policy in Table 2.3: 6.801, 8.143, and 13.425, physical capital accumulation is reduced by the policy in both the short and long run. Again, this follows from the decreased incentive to save, and most of this is due to the higher tax rate.

C_t jumps slightly down from 1.129 to 1.111 on the new saddle path toward the post-policy equilibrium because of households' forward-looking behavior in the face of the unexpected announcement of the policy. It then increases to 1.129 after one year, 1.135 after ten years, and 2.100 after thirty-five years. Comparing to its counterparts without the policy in Table 2.3: 1.152, 1.379, and 2.274, consumption is hurt by the policy in both the short and long term.

Relative to the pre-policy equilibrium, the percentage change of the discounted utility summed over time after the policy is -13.12% right after the policy shock, which is an immediate pain. This pain is quickly relieved since the percentage change becomes smaller at −0.39% after one year. It then reverses the sign and reaches +54.09% after ten years, and +73.10% after thirty-five years. Hence, under the large parameter set, the short-term pain is more than offset by welfare gains in the medium and long term.

2.6.3 Post-Policy Equilibrium and Transitional Dynamics Using Small Parameter Set

Table 2.5 shows the results using the small parameter set. H_t increases from 1 to 1.007 after ten years, to 1.016 after thirty-five years, and eventually converge to 1.021.

Table 2.5: Equilibrium shift and transitional dynamics caused by the policy that reallocates labor from manufacturing to health sector to double health labor from 5% to 10% of the nation's workforce using the **small** parameter set

	Pre-policy equilibrium	Post-policy dynamics					
	$t = 0$	$t = j$	$t = 1$	$t = 10$	$t = 35$	$t = 50$	$t = \infty$
L_t	5%	10%	10%	10%	10%	10%	10%
H_t	1	1	1.001	1.007	1.016	1.018	1.021
x_t	16.94%	16.55%	16.55%	16.55%	16.55%	16.56%	16.56%
A_t	0.3	0.289	0.289	0.289	0.290	0.290	0.290
τ_t	0.3	33.07%	33.07%	33.07%	33.07%	33.07%	33.07%
r_t	6%	5.65%	5.65%	5.65%	5.65%	5.65%	5.65%
g_t	2%	1.58%	1.58%	1.58%	1.58%	1.58%	1.58%
K_t	6.667	6.667	6.773	7.807	11.597	14.707	∞
C_t	1.129	1.103	1.121	1.292	1.920	2.435	∞

Notes: Note: j denotes jump, which means the post-policy initial value.

x_t jumps down from 16.94% to 16.55%, then almost remains at that level throughout the time. A_t immediately decreases down from 0.3 to 0.289 and then slightly increases throughout the time and converges to 0.290, still lower than 0.300 when under no policy. τ_t immediately jumps up to 33.07% and remains at this level all the time. r_t immediately drops from 6% to 5.65%, then stays at this level throughout the time, lower than that without the policy. K_t rises from 6.667 to 6.773 after one year, 7.807 after ten years, and 11.597 after thirty-five years of introducing the policy. C_t jumps from 1.129 down to the new initial value 1.103 on the new saddle path. It then increases to 1.292 after ten years and 1.920 after thirty-five years.

Comparing with the counterparts using the large parameter set in Table 2.4, H_t grows much slower. x_t falls much more. A_t is lower throughout time which suggests that when health parameters (in particular, the productivity-enhancing effect of health β) are all small, productivity increases very slowly. τ_t is negligibly higher. r_t becomes even lower. K_t decreases more. C_t drops more in the short term and increases less in the medium and long term.

More importantly, g_t immediately drops down to 1.58%, and remains at this level and fails to grow back throughout the time horizon. Comparing with its counterparts — 1.81% after half of a century and 1.85% at the post-policy equilibrium in Table 2.4, the economic growth is hurt by investing in health regardless of whether we use the large or the small parameter set. But it is hurt much more when health parameters are all small than they are all large.

Regarding the welfare effect, relative to the pre-policy equilibrium, the percentage change of the discounted utility summed over time after policy is -19.11% right after the shock, -17.71% after one year, -11.86% after ten years, and -10.32% after thirty-five years. This suggests that when all health parameters are small, the health investment makes households worse off. That is, there is no gain throughout the time.

2.6.4 Sensitivity Tests of ξ, β, and γ

We now report how the policy's welfare effect is sensitive to the values of the three parameters ξ, β, and γ. We begin with the households' value of health versus material consumption (ξ). Table 2.6 reports the percentage changes of discounted utility

Table 2.6: Sensitivity test of the weight (ξ) of health relative to consumption in utility by reporting the percentage change of discounted utility summed over time with policy relative to that without policy

	Post-policy dynamics		
	$t = 1$	$t = 10$	$t = 35$
Large set $\beta = 0.95$, $\gamma = 0.09$			
$\xi = 14$	-0.39%	$+54.09\%$	$+73.10\%$
$\xi = 5$	-8.93%	$+9.71\%$	$+18.15\%$
Small set $\beta = 0.05$, $\gamma = 0.03$			
$\xi = 14$	-14.92%	$+2.70\%$	$+7.82\%$
$\xi = 5$	-17.71%	-11.86%	-10.32%
Middle set $\beta = 0.50$, $\gamma = 0.09$			
$\xi = 14$	-2.97%	$+51.70\%$	$+69.21\%$
$\xi = 5$	-11.51%	$+7.32\%$	$+14.27\%$
Middle set $\beta = 0.50$, $\gamma = 0.03$			
$\xi = 14$	-14.03%	$+3.50\%$	$+9.09\%$
$\xi = 5$	-16.82%	-11.06%	-9.05%

summed over time with the policy relative to that without the policy, when the value of health is large ($\xi = 14$) and when it is small ($\xi = 4$), for different combinations of the other two parameters. Comparing the two rows within each panel, that is, fixing β and γ values, the size of ξ affects the welfare changes considerably both in size and sometimes even in sign. For example, comparing the first and second rows in the first panel, that is, fixing $\beta = 0.95$ and $\gamma = 0.09$, the welfare gains when health's weight is 14 are 4–6 times higher than the gains when health's weight is 5.

Table 2.7 reports the sensitivity tests of β and γ by reporting the percentage change of discounted utility summed over time. To detect whether the size of the

Table 2.7: Sensitivity test of return to labor in health production (γ) and return to health in manufacturing productivity (β) by reporting the percentage change of discounted utility summed over time with policy relative to that without policy

	Post-policy dynamics		
	$t = 1$	$t = 10$	$t = 35$
	$\xi = 14$, $\gamma = 0.09$		
$\beta = 0.95$	-0.39%	$+54.09\%$	$+73.10\%$
$\beta = 0.50$	-2.97%	$+51.70\%$	$+69.21\%$
$\beta = 0.05$	-5.75%	$+49.27\%$	$+65.36\%$
	$\xi = 5$, $\gamma = 0.03$		
$\beta = 0.95$	-15.95%	-10.27%	-7.78%
$\beta = 0.50$	-16.82%	-11.06%	-9.05%
$\beta = 0.05$	-17.71%	-11.86%	-10.32%
	$\xi = 5$, $\beta = 0.05$		
$\gamma = 0.09$	-14.29%	$+4.89\%$	$+10.41\%$
$\gamma = 0.03$	-17.71%	-11.86%	-10.32%
	$\xi = 14$, $\beta = 0.50$		
$\gamma = 0.09$	-2.97%	$+50.70\%$	$+69.21\%$
$\gamma = 0.03$	-14.03%	$+3.50\%$	$+9.09\%$
	$\xi = 5$, $\beta = 0.50$		
$\gamma = 0.09$	-11.51%	$+7.32\%$	$+14.27\%$
$\gamma = 0.03$	-16.82%	-11.06%	-9.05%
	$\xi = 14$, $\beta = 0.95$		
$\gamma = 0.09$	-0.39%	$+54.09\%$	$+73.10\%$
$\gamma = 0.03$	-13.16%	$+4.29\%$	$+10.36\%$
	$\xi = 5$, $\beta = 0.95$		
$\gamma = 0.09$	-8.93%	$+9.71\%$	$+18.15\%$
$\gamma = 0.03$	-15.95%	-10.27%	-7.78%

productivity-enhancing effect of health (β) matters in the welfare effect of the policy, we focus within the first and second panels respectively. When fixing $\gamma = 0.09$ and $\xi = 14$, the welfare gain after thirty-five years of policy is +73.10% when β is 0.95, +69.21% when β is 0.50, and +65.36% when β is 0.05. These values are not substantially different from one another. Similar patterns hold when fixing $\gamma = 0.03$

and $\xi = 5$ except that all signs reverse. Thus, the signs and sizes of the welfare changes do not obviously differ among values in the entire range of β. This suggests that the contribution of heath's productivity-enhancing effect to the policy's welfare effect is limited.

To detect whether the effectiveness of labor in producing health (γ) matters for the welfare effect of the policy, we compare within the third – seventh panels respectively. The third panel fixes $\beta = 0.05$ and $\xi = 5$, the welfare gain after thirty-five years of the policy is +10.41% when γ is large (0.09). But it turns to a welfare loss of −10.32% when γ is small (0.03). Similarly, the welfare gain when γ is large reverses to welfare loss whenever γ is small, while fixing ξ small and β moderate (in the fifth panel) and large (in the seventh panel). Whereas, the fourth panel fixes $\beta = 0.50$ and $\xi = 14$, the policy results in welfare gains regardless of the size of γ. But the welfare gains are 8–16 times higher when γ is big (0.09) than the gains when γ is small (0.03). A similar result is observed in the sixth panel that fixes $\beta = 0.95$ and $\xi = 14$.

Therefore, the sensitivity tests in Tables 2.6 and 2.7 indicate three results. First, whenever β and γ are fixed, the size of health's relative weight (ξ) substantially affects the size or sign of the welfare effect of the policy. Second, no matter how large or small the effectiveness of labor in health production (γ) and the weight of health in utility (ξ) are, the size of the productivity-enhancing effect of health (β) does not substantially affect the welfare effect of the health investment policy. Third, no matter how large or small the productivity-enhancing effect of health (β) and the weight of health in utility (ξ) are, the magnitude of the effectiveness of labor in health production (γ) affects the welfare effect of the policy substantially, and it affects even the sign of the welfare effect when ξ is small. This suggests that, regardless of the relative values of health and consumption in utility, as long as the health investment has a sufficiently large return in generating health status, it generates welfare gains.

2.6.5 A Simulation Suitable for Canada

Finally, Table 2.8 shows the simulation using a set of parameter values that are particularly suitable for Canada. In this stylized simulation, γ is as large as 0.09 calculated from Canadian historical data, while β is as small as 0.05 since the aged population who need health services most are less likely to participate in the workforce. In both rows of different ξ values, the negative signs of the percentage change in utility immediately and one year after the policy reverse to positive signs in the medium run. This means that no matter how big or small the health's relative weight in utility is, the health investment policy always causes a short-term pain but a medium- and long-term gain in households' well-being. In particular, the welfare gains when ξ is large (14) are about 6–10 times higher than the gains when ξ is small (5) in the medium and long run. This suggests that no matter whether we believe that households value health much higher than consumption, the investment in health always generates welfare gains. However, it also lowers the long-run economic growth rate to 1.59% at the post-policy equilibrium.

Table 2.8: A simulation using parameter values particularly suitable for Canada by reporting the percentage change of discounted utility summed over time with policy relative to that without policy

	Post-policy dynamics		
	$t = 1$	$t = 10$	$t = 35$
	$\beta = 0.05$, $\gamma = 0.09$		
$\xi = 14$	-5.75%	$+49.27\%$	$+65.36\%$
$\xi = 5$	-14.29%	$+4.89\%$	$+10.41\%$

2.7 Conclusions and Policy Implications

This chapter examines the competing effects of an investment in health on the welfare and the endogenous economic growth rate in a two-sector economy. It separates the health and non-health firms' production functions and includes health status in households' utility function. It explores, both analytically and numerically, the full equilibrium and transitional dynamics after a once-for-all health investment policy that reallocates labor from the manufacturing to the health sector. It captures some essential trade-offs associated with health investments. The sensitivity tests help us to verify how robust our findings are to alternative calibrations of health parameters such as the households' value of health versus material consumption, the effectiveness of labor in health production, and the productivity-enhancing effect of health. Thus, the sensitivity test of the productivity-enhancing effect of health helps us to address the challenges faced by both developed and developing countries with publicly-financed health sectors.

Our model has limitations, and it can be extended in several directions. We have assumed that the non-health governmental spending does not appear in the households' utility function. But if the health investment policy slowed economic growth and the share (z) of governmental programs' expenditure of the total output were fixed as we have assumed, governmental spending on the non-health products would be reduced endogenously. Hence, households' living standard would be harmed. However, our assumption can be justified when the essential trade-off in question does not involve the governmental spending on the non-health programs.

We have assumed that all households are identical. Hence, we cannot discuss the relationship between inequality and economic growth in this model which would be interesting as emphasized by Aghion *et al.* (1999). If we had a poor group that was dependent on the government for transfer payments, the lower economic growth rate and hence the smaller government's tax base would be a concern since it would limit the government's ability to pursue income redistribution.

We have also assumed that the health investment policy is unexpected and once-for-all. Thus, as pictured in Figure 2.1, the economy jumps to the new saddle path immediately and sharply after policy. However, if policies were preannounced and hence expected, or imposed incrementally, the equilibrium would be shifted gradually. This features more complicated transitional dynamics that would require further research.

The model can be modified to eliminate the use of linear approximations of all relationships, and this may have important effects. Turnovsky *et al.* (2010) find that in a one-sector endogenous growth model where governmental spending enters the production function, the reliability of a linear approximation depends on the nature of governmental spending and the size of the adjustment speed. If the productive governmental spending were an additional state variable as would be the case if that spending were focused on an accumulating stock such as infrastructure, then the error of linearly approximated transitional dynamics can be big. This is the case in our paper since health enters the manufacturing production as a state variable, so their analysis does raise a concern. Further, they find that the smaller the adjustment speed per annum, the bigger the linear approximation error. This is also a concern for our analysis since our annual adjustment speed is merely 4%. So there is a scope for further fruitful analysis using non-linear approximation or even non-linear dynamic system. For instance, Fanti *et al.* (2014) specifies a non-linear dynamic system where health investment is financed through an income tax, and this investment raises people's survival probability by the built-in random mechanism that the young receive unpredictable bequests from the old. They find that a deterministic chaos rather than equilibrium occurs when the wage income tax rate that government levies falls is large. Thus, they warn countries of economic chaos due to a big investment in health. But it must be remembered that virtually no macroeconomic model of health that is accessible for policy makers has considered transitional dynamics. So our analysis of transition paths is an important stepping stone for future analysis of non-linear dynamics involving health investment.

Given our policy focus, our model is designed to be relatively simple while capturing some essential trade-offs in order to provide important insights into a range of issues associated with the investment in health for policy makers. We offer four major conclusions and policy implications. First, a policy that invests in health (here modeled by a reallocation of labor from the manufacturing to the health sector) improves health status hence welfare in the long run, but harms the economic growth in both short and long term. The relative sizes of these competing effects depend on the specific health features of the country. When the productivity-enhancing effect of health and the effectiveness of labor in health production are both large enough, more than half of the immediate harm in economic growth recovers in the long run. In other words, the economic slowdown is partially offset in the long run by the enhanced productivity.

Such long-run welfare and growth effects we find are consistent with existing theoretical literature, in particular, with Van Zon and Muysken (2005) (p.58–60) who find from a numerical analysis that at the equilibrium "the underachievment with respect to the provision of health services has a direct negative effect on welfare"; and that "the growth of the economy with an aging population will be affected in an unambiguously negative way." Our finding of a definite negative growth effect contrasts with Gong *et al.* (2012)'s findings in which the impacts of health investment on growth, both theoretically and empirically, are ambiguous. Therefore, our findings together with relevant literature challenge the policy rationales of World Bank (1993) and World Health Organization (2001) in the sense that good health, though improves welfare, increases neither substantially the economic productivity of workers nor the economic growth rate of countries.

Second, no matter whether we believe that households value health much higher than material consumption, households are always better off by the health investment policy as long as the effectiveness of health investment in producing health is large. Households can be worse off by such a policy only if both health investment is not productive in generating health and households value health relatively lower (5 times and less) than consumption. In the real world at the disaggregate level, there can be many kinds of particular investments in health from the health sector such as health professionals, primary care, hospital beds, and diagnostic tests, and from the non-health sectors such as public funding, improved housing, education, transportation and other heath-affecting services. Policy makers need to be aware of such a sensitivity of the finding to the effectiveness of health investments in producing health, so that they can invest wisely to achieve welfare gains instead of pains.

Third, this chapter's emphasis on investment in the public health care sector applies to developed countries who have a publicly funded health sector like Canada, the United Kingdom, and Scandinavian countries. For these countries where the productivity-enhancing effect of health is small but the effectiveness of labor in health production is still large, an investment in health is likely to lead to a welfare gain for households over the medium and long run. However, such an investment is likely to slow down economic growth in both the short and long run. Such a lowered economic growth shrinks the size of the pie of national income that would have been larger (than without the policy) for the government to redistribute to the poor and reduce the poverty. Therefore, only when these countries care about welfare more than economic development *per se*, the substantial long-term gains in health and welfare are worth the short-term small pain in consumption and long-term moderate harm in economic growth.

Fourth, our model serves as a sensitivity test for Van Zon and Muysken (2005) since it constructs a health production function traditionally documented in macroeconomic literature rather than population-based function as done in their model. Our model is also a sensitivity test for Hall and Jones (2007) in three ways. It does not rely on a strong interaction between health and consumption in households' utility formulation as done in their basic model. It assumes Ricardian equivalence (the young is born to replace the old once they die, and the welfare of the young is of equal concern for the "family dynasty") as a special case for their age-specific overlapping generations. It tests the sensitivity of conclusions to the households' much larger value of health than consumption, as assumed in their full model. Therefore, given our conclusion about the welfare gain of health investment as long as health labor is productive enough regardless of the relative weight of households' value on health, our model shows the robustness of the Van Zon and Muysken (2005) and Hall and Jones (2007)' findings.

Appendix

Alternative calibration of the effectiveness of labor in health production (γ) using Van Zon and Muysken (2005):

Another way to find the γ value follows from the analysis in Van Zon and Muysken (2005), which can be explained by referring to their Figure 2.1 (our Figure 2.4):

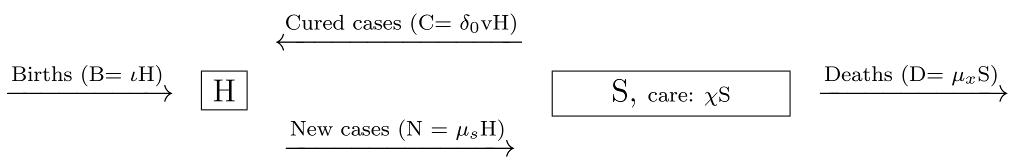


Figure 2.4: Flow chart of the healthy and sick population from Fig. 2.1 Van Zon and Muysken (2005)

This is a flow chart similar to the commonly-used Susceptibles-Recovery model in mathematical epidemiology (Brauer *et al.* (2008)). In explaining this model, we use the notation from Van Zon and Muysken (2005), so the variables differ from how we have defined these symbols in our model. The total population (P) is divided into two groups, the healthy (H) and the sick (S). A proportion (ι) of the healthy people give new births (B) to the healthy pool. Whereas among the sick, some (C) are cured back to full health, as a proportion (δ_0) of health labor which is further a proportion (v) of the healthy population. Some as a proportion (χ) of the sick (S) are currently cared by the other type of health labor (u) hence stay in sick. Others (D) die as a proportion (μ_x) of the sick. Besides, new cases (N) come from the healthy population as a proportion (μ_s) to the sick population. Thus, the dynamic system of the healthy and the sick can be summarized as below

$$\dot{H} = B + C - N = \iota H + \delta_0 v H - \mu_s H, \tag{2.41}$$

$$\dot{S} = N - C - D = \mu_s H - \delta_0 v H - \mu_x H. \tag{2.42}$$

These two equations are useful, because the goal is to find the steady-state population health h^* (that is, the steady-state proportion of the healthy population in the total population) at the equilibrium $\dot{h}/h = \dot{H}/H - \dot{P}/P = 0$. On the one hand, we can get $\dot{H}/H$ by dividing equation (2.41) by H. On the other hand, recall the total population P is the sum of the healthy and the sick, this implies $\dot{P} = \dot{H} + \dot{S}$ whose expression is obtained by substituting equations (2.41) and (2.42). By solving the resulting $\dot{h}/h = 0$ for h^*, we obtain equation (2.43) where the steady-state population health positively depends on the health labor who cure (v^*) sick people back to healthy.

We now turn to health labor who care (u) the sick people remaining in illness or dying. Such a demand for care services equals the supply of care services at the equilibrium; that is, $u^* h^* P = \chi(1 - h^*)P$. By solving this equation for u^*, we obtain equation (2.44). It states that the steady-state population health can also negatively depend on the health labor who care, which indicates an interesting trade-off between the curing (v^*) and caring (u^*) health labor. Finally, the curing and caring health labor are subject to the constraint of total health labor L in equation (2.45).

To summarize, Van Zon and Muysken (2005) have derived three equations of three endogenous variables h^*, v^*, and u^* from the system at the equilibrium

$$h^* = \frac{\delta_0}{\tau + \mu_x} v^* + (1 - \frac{\mu_s}{\tau + \mu_x}), \tag{2.43}$$

$$u^* = \chi \frac{1 - h^*}{h^*}, \tag{2.44}$$

$$L = v^* + u^*. \tag{2.45}$$

We can solve the three equations for

$$\frac{dH}{dL} = \frac{\frac{\delta_0}{\tau + \mu_x}}{1 - \frac{\chi \delta_0}{(\tau + \mu_x)(h^*)^2}},$$

which is equivalent to $dH/dL = \gamma H/L$ in our model. By subsituting the assumed parameter values in Van Zon (2005) Table 2.2, we obtain γ as 0.03.

Chapter 3

Forecasting Health Expenditure: Methods and Applications to International Databases

3.1 Introduction

Health expenditures have been increasing in the past decades among developed and developing countries (Newhouse, 1977; Culyer, 1988; Gerdtham *et al.*, 1992; Panopoulou and Pantelidis, 2011; Lago-Peñas *et al.*, 2013; Lorenzoni *et al.*, 2014). This raises global concerns about health and the economy, and calls for fiscal preparations and policy interventions to ensure future fiscal sustainability (European Union, 2009; Centers for Medicare & Medicaid Services, 2011). Forecasting health expenditures is crucial for policy applications required by governmental organizations and central banks. For example, under-predicting public health expenditures can result in unmet health needs that are eligible for public support, or in shortages of infrastructure investments in hospitals and health human resources. Excess public health expenditure on hospitals and physician services crowds out financial resources that might have been allocated to other equally important sectors such as education for human capital accumulation. It may also absorb labor that could have been directed to the final goods sector and hence reduce manufacturing outputs, which in turn may hurt both welfare and economic growth over time.

This chapter primarily addresses the methodological question: how can we best forecast health expenditure in a systematic way? It then answers the empirical question: given our best estimates, how much would worldwide health expenditures be in the short- and medium-run future (e.g., 2015-2025)? This chapter examines a comprehensive set of measurements and projection models for forecasting health expenditures, and carefully tests their forecast performance based on formal criteria and by using recent and comparable data provided by international sources. The methods and empirical results of health expenditure forecasts can inform the policy making process by projecting needed funds and identifying gaps between the amounts of monetary resources needed and those available.

Literature relevant to these questions involves two sub-fields of economics: health economics (esp., expenditure), and economic forecasting. The majority of the health expenditure literature tries only to understand past drivers of health expenditure, rather than project health spending into the future. Although we are aware that there are hundreds of such past-estimation articles using micro data, we only focus on those using aggregate data since our data is at the country level. The seminal paper by Newhouse (1977) regressed per capita health expenditure on per capita GDP using cross-sectional data from OECD countries, and found that GDP is a statistically and economically significant factor in explaining health spending. Empirical studies using the OECD database have tried various econometric models and estimation methods. Gerdtham and Jonsson (2000) divided studies into two generations. The first-generation studies used cross-sectional data for a single year (or selected years) to examine within-country determinants of, and cross-country differences in, health expenditure. While these studies paid attention to methodological issues such as the appropriate choice of the currency unit (or conversion factors) (Parkin *et al.*, 1987, 1989), they have been criticized owing to shortcomings inherent in cross-sectional data estimates. For example, small sample size and omission of the country- and time-invariant variables can generate inconsistent estimates of the regression coefficients. These criticisms were addressed by the second-generation studies that used panels of countries, where a relatively long time series of annual data was available for each country. These latter studies, however, faced other methodological issues such as non-stationarity[1], cointegration[2] between two and more variables, and heterogeneity across countries (Gerdtham *et al.*, 1992; Barros, 1998).

In addition, the broader economic literature on forecasting emphasizes fundamental methodological challenges as well. First, forecasting imposes an assumption of continuity (NBER, 1966; Theil, 1966) — that the pattern of data in the past persists in the future. Second, an explanatory forecast ignores the unquantifiability of certain variables such as technological progress. Third, forecasting has to acknowledge the "unknown uncertainty" (Clements and Hendry, 2004, 2011) that we have not discovered and hence are unable to incorporate the true data generating process. Rigorous forecasts, however, can be achieved from sound economic theories and econometric methods, and careful assessment of the performance of empirical models according to three formal criteria — accuracy, precision, and known uncertainty (Clements and Hendry, 2004, 2011). Forecasts based on sound grounds do add new information about the status of the future, hence can stimulate governmental reactions within a health system (Box *et al.*, 2008). If a forecast depicts a gloomy picture ahead, then policy makers can respond by changing certain conditions to prevent its realization. Rigorous forecasts, better than *ad hoc* guesses, can successfully guide policy makers to enhance the likelihood of a favorable outcome and avoid an undesirable outcome.

Few studies were found to integrate health expenditure analysis and economic forecasting by searching truncated terms "health expenditure AND (project OR forecast)" in two commonly used economic literature databases (EconLit, IDEAS). To

[1]Non-stationarity is defined in the literature in various but similar ways: "Data in economics is often nonstationary, namely, has changing means and variances over time. Failure to allow for such specific characteristics may result in inferior forecasts of aspects of interest " (Clements and Hendry, 2004) (p.8). "Stationarity means that . . . the data fluctuate around a constant mean, independent of time, and the variance of the fluctuation remains essentially constant over time " (Makridakis *et al.*, 1998) (p.324).

[2]Cointegration intuitively means that a long-run relationship exists between variables.

forecast provincial health expenditure, Di Matteo (2010) used measures of real per capita Canadian provincial governmental spending on health care from 1965 to 2008, economic and demographic regressors such as provincial GDP, provincial population in total and the proportion of the elderly, federal cash transfers, and provincial governmental revenues and expenditures. The study used classical regression estimators including ordinary (OLS) and generalized least squares (GLS), focused only on regressors with statistically significant coefficient estimates, and used simple extrapolations of their respective historical growth rates into the future to generate future values of health expenditure. The study neither included specification tests for the consistency of the pooled OLS and GLS estimators (e.g., whether the constant coefficients α and β are appropriate for the purpose of using pooled OLS estimator) nor formally reported forecast performance of the models.

Getzen and Poullier (1992) conducted forecasts of national health expenditure using data for 19 OECD countries. They used estimates based on 1965–1979 data to forecast within the sample (1980-1987) using the naïve method, exponential smoothing, and multivariate (inflation and GDP) models respectively for each country and the pooled-country panel. They then measured the mean absolute error (MAE), an indicator of forecast accuracy, and compared MAE of each method with that of the naïve method. For each country, the combination or average of forecasts obtained from single forecasting models was found to be more accurate than any single model alone. However, the finding that multivariate regressions offered more accurate forecasts than time series models such as ARIMA is inconsistent with much of the literature on economic forecasting (Armstrong, 2001; Makridakis *et al.*, 1998). Such a finding may arise because, as has been recognized, an insufficient number of observations leads to unstable estimated parameters. Overcoming this limitation requires waiting for more data to accumulate. This is the case in the present chapter wherein an additional 20 years of data are used to forecast.

Moreover, Getzen (2006, 2007) has argued that model specifications of health expenditures vary with temporal and spatial dimensions, that is, the chosen length of time horizon and breadth of the observational unit. He presents a framework to categorize health expenditure projection methods using conceptual metrics that consists of four aspects: observational units at the macro- and micro-levels; time span over the short, medium, and long run; measures of Total Health Expenditure (THE); and incorporation of various regressors. The best indicator of the short-run growth of nominal THE was its growth rate in the previous year, plus a time trend, and a few regressors including employment and inflation rates, rather than other factors. Endogenous and pre-determined variables such as health, demographics (e.g., aging), hospital infrastructure, and physician supply change little in the short run (i.e., 1–3 years), and other factors have already been constrained within the existing budget (e.g., funds for research and development, affordability of technology adoption). Therefore, their effects on THE neutralize overall, and their attributions to the short-term forecast of THE are argued to be negligible.

Good indicators of medium-run (i.e., 3 –10 years) forecasts of THE can be achieved using only past values of THE, plus one and only one crucial regressor — national income — both measured in real, per capita terms. This approach was adopted for this chapter for medium-run forecasts. We are not interested in the long run since anything can occur over such a long period.

As Astolfi *et al.* (2012) point out, at the macro (e.g., country) level, there is another branch of forecasting that uses structural models — computable general equilibrium (CGE) models. These models rely on economic theory, which is helpful in selecting variables as potential drivers of HE and in imposing assumptions (e.g., existence and number of equilibria in CGE models) that may decrease parameters' estimation error (Elliott and Timmermann, 2008). This type of model is particularly helpful to answer "what if" questions regarding the results of exogenous policy interventions. In contrast, because they do not specify mechanisms for policy and changes in both health and non-health sectors, forecasts based on reduced-form models cannot clearly explain "why" and "how" an effect occurs. However, the CGE approach is both more difficult and more costly because it requires the construction of formal structural models of health and the economy. Constructing such a structural model is beyond the scope of this thesis chapter.

Besides forecasts from academia, governmental agencies also provided forecasts. For instance, the U.S. Medicare Trustees provide detailed short-term projections of health sub-sectors, and long-run forecasts of total Medicare costs. They first project the percentage growth rate of real per capita GDP, which is then used to project aging-related Medicare costs (Centers for Medicare & Medicaid Services, 1991). They then add 1% to incorporate a technology effect (Centers for Medicare & Medicaid Services, 2000, 2004). This has been criticized as arbitrarily adding an excess growth rate of GDP (Getzen, 2007). Without being tested, aging was deemed by the OECD (2003) and the European Union (2008) as a major determinant of health care costs and an attribute to forecasts of health expenditures. It has been argued that aging *per se* is neither a speedy nor a substantial driver of health care cost growth, compared with the effects of growth of the national income and budget (Denton *et al.*, 2002; Evans *et al.*, 2001). Therefore, it is useful to forecast health expenditures both excluding (e.g., in univariate time series models) and including (e.g., in panel data models) demographic variables separately.

To sum up, Getzen and Poullier (1992) obtained an unusual result — that panel data models perform better than time-series models — and argued that people should examine this issue as more data accumulate. Our paper does exactly this and obtains the more conventional result, in a sense, overturning Getzen and Poullier's initial finding. The comprehensive literature review conducted by Astolfi *et al.* (2012) suggested that with better computing power and more refined data, the future of forecasting is complicated micro-level models. The objective of simply getting accurate forecasts of future spending from modeling, however, needs to be carefully distinguished from that of understanding the determinants of expenditure growth. Modeling and understanding the determinants of expenditure growth (whether using micro data or CGE macro models) require considerably more data and effort, and may still do worse for pure forecasting. Our paper confirms this. At the same time, our paper contributes to the call (articulated in Astolfi *et al.* (2012)) for more rigorous methods, for more transparency, and for better assessment of forecast performance.

3.2 Specification of Projection Models

Projection models require three basic assumptions: past information is available, it is quantifiable, and certain characteristics of the pattern of such data continue into the future, i.e., the assumption of continuity or constancy (NBER, 1966; Theil, 1966). A short-term forecast usually spans 1 to 3 years, whereas intermediate and long-range predictions are typically 3 to 10 years and beyond 10 years, respectively (Getzen, 2000). Generally, the projection model in simple logarithmic form is

$$lnY_t^f = X_t b \quad \text{(t=T+1,T+2,\ldots,T+F)}, \tag{3.1}$$

where T means the time horizon of the estimated past data; F refers to the number of forecast periods; Y is the dependent variable that is to be forecasted at a point in time as Y_t^f; and X is the independent variable that is used to forecast Y. E($ln(Y_t^f)) = X_t$b. Var($ln(Y_t^f)) = X_t$[Var(b)]X_t'. The forecast error $e_t = ln(Y_t)$-$ln(Y_t^f)$ (t=T+1,T+2,...). E$(e_t) = 0$ and Var$(e_t) = s_t^2$ where s is the estimated standard error. Given a level of significance $\alpha > 0$, the confidence interval of forecast of $ln(Y_t)$ is defined using

$$\Pr\left[lnY_t^f - t_{\frac{\alpha}{2}} s_t \leq lnY_t \leq lnY_t^f + t_{\frac{\alpha}{2}} s_t\right] = 1 - \alpha.$$

Based on the taxonomy of projection models presented by Clements and Hendry (2004), we categorize projection methods according to the following aspects: time span (short vs. medium vs. long run); data pooling (single- vs. pooled-country); type of model (statistical smoothing vs. econometric time series data vs. panel data); number of independent variables (univariate vs. multivariate); existence of causality (non-causal vs. causal[3]); existence and number of time lags (static vs. dynamic); and directness of forecast procedure (direct vs. combined). Based on meaningful combinations of these aspects, we divide six commonly used reduced-form models into three time series and three panel data models (Table 3.1). Each model features certain of the above-listed categories. As we consider the six projection models below,

[3]Admittedly theorists who favor structural models may be less likely to agree with empiricists′ causal interpretation of certain reduced-form models.

each is found to overcome certain methodological weaknesses of the preceding model.

Table 3.1: Projection models

Model class	(Estimation) Model specifications
Time series data	
M1. Exponential smoothing	$y_t^f = y_{t-1}^f + \alpha(y_{t-1} - y_{t-1}^f)$
M2. ARIMA(p,I,q)	$y_t = b_0 + b_1 y_{t-1} + \ldots + b_p y_{t-p} + e_t - d_1 e_{t-1} - \ldots - d_q e_{t-q}$
M3. VAR	$Y_t = A_0 + A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + e_t$
Panel data	
M4. Static, fix effects	$y_{i,t} = a_{i,fe} + \gamma_1 x_{1i,t} + \ldots + \gamma_n x_{ni,t} + e_{i,t}$
M5. Dynamic	$\Delta y_{i,t} = a + b_1 \Delta y_{i,t-1} + \ldots + b_p \Delta y_{i,t-p} + \gamma_1 x_{1i,t} + \ldots + \gamma_n x_{ni,t} + \Delta e_{i,t}$
M6. Nonstationary & cointegrated	$y_{i,t} = \gamma x_{i,t} + e_{i,t}$, where $x_{i,t} = x_{i,t-1} + u_{i,t}$

Notes: f means forecast. α is the weight on previous forecast error. Y refers to the matrix of dependent variable y and independent variable x. A is the matrix of coefficients. Δ represents first difference. $y_{i,t}$ in M6 is cointegrated with $x_{i,t}$, and $x_{i,t}$ is integrated process of order 1 for all i. Letters e and u refer to different error terms. Parameter p means time lag of a variable. q is the time lag of error. γ is the coefficient of a variable. Estimated values of the same parameter are not necessarily the same in differently specified models.

The exponential smoothing model (M1) requires only a single series. Therefore, it produces forecasts with little information at low costs. It aims to reduce randomness and estimate the trend component of a series. Observations are given weights that exponentially decline as observations become remote (Holden *et al.*, 1990). Both double exponential (DE) smoothing and its alternative Holt-Winters non-seasonal smoothing are conducted to project the series mean (and hence the point forecast) rather than the variance (and hence the forecast interval), since the latter becomes invalid when the assumption of independently-and-identically-distributed errors (i.i.d.) does not hold. The autoregressive integrated moving average model ($AR_pI_dMA_q$, M2) is argued by Granger (1989) to improve forecasts with low marginal cost since the information it requires is still about a single variable and with high marginal benefit as one more recent (rather than average) observation gives better information for future forecasts. To confirm estimates of the number of Y′s own lagged values, p, and the number of current and lagged residuals, q, we use Durbin-Watson, Durbin′s alternative, and Breusch-Godfrey tests for lower- and higher-order serial correlations of residuals. To diagnose d, the number of times of differencing, it is necessary to difference the series into stationarity. We use the Augmented Dickey-Fuller (ADF) test and other two tests for unit roots — the Philips-Perron (Philips and Perron, 1988) and DF-GLS (Elliott *et al.*, 1996) tests. ADF imposes the strict assumption that errors are serially uncorrelated and homogeneously distributed, while Phillips-Perron allows errors to be weakly dependent and heterogeneously distributed. The DF-GLS test successfully distinguishes effects of unit roots from the deterministic trend and hence has greater power than ADF.

The vector autoregression (VAR, M3) model allows for endogeneity and bi-directional correlation (not necessarily causality (Sims, 1972)). The cointegration between variables (e.g., THE and GDP) signifies further possible cointegration between subsets of these variables (Dolado *et al.*, 1990), which justifies our examination of the relationship between public health expenditure (PHE) and general governmental revenue (GGR). Multicollinearity is not a concern as long as the assumption of continuity (i.e., the pattern of such multicollinearity continues into the future) holds. All series are checked for cointegration by applying ADF, Philips-Perron, and DF-GLS tests to the residuals. The stability of VAR and the normality of disturbances distribution are also tested. The model that passes all specification tests is used for estimation, point forecast, and forecast interval. If more than one adequate specification occurs, we select the better-estimated one according to the principle of parsimony using minimum information criteria (Akaike′s, AIC; Bayesian, BIC) for lags and the three criteria of forecast performance (introduced in the next section). We do not estimate the Vector-Error-Components (VEC) model because for some countries the GDP (or GGR) is not cointegrated with THE (or PHE) through the Johansen's cointegration test. This justifies using VAR generally and consistently for all countries rather than VEC conditional on cointegration.

Compared to time-series data models, panel data offers three advantages: a larger number of observations which enhances estimation precision, the possibility of establishing causation under the assumptions of the fixed effects (FE) model (Cameron and Trivedi, 2010), and the ability to estimate country-specific effects. However, both the number of countries N and the number of years T of the series at the annual national level are relatively small, which does not allow formal tests for poolability and country-specific effects (Cameron and Trivedi, 2010). We specify a constant slope β, as is commonly done in the literature when using such aggregate data, and use the Hausman test to indirectly test the consistency of Random Effects (RE) by comparing it with FE. Other estimators are disregarded because the "within" estimator is equivalent to FE if the intercepts are fixed effects and errors are i.i.d., while "pooled" and "between" estimates are similarly consistent with and usually less efficient than those obtained from RE (Cameron and Trivedi, 2010). We find that RE is rejected by the Hausman test and use panel-robust standard errors that correct both serial correlation and heteroskedasticity.

Such static models, however, impose a strong assumption of exogeneity[4] that can be relaxed by dynamic models. The latter further extract the state dependence of a country — as distinct from unobservable heterogeneity across countries — from the observed total variations of a country. For instance, let y and x be THE and GDP respectively. If FE were true, then true state dependence would attribute the currently high level of THE in a given country to its own high historical level, after controlling for regressor GDP, rather than to the significant difference of this country from others.

[4]The errors have mean zero conditional on the past, current, and future values of regressors.

Therefore, dynamic models can explain state dependence that static models cannot. If state dependence were true for a country, then dynamic models would suggest a different corresponding policy that focuses on the historical problem of this country only. We apply estimators proposed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). These country-panel models, similar to time-series data models, confront the nonstationarity problem since they incorporate the time series component in pooled cross-sections. We diagnose such nonstationarity and cointegration, and estimate using cointegrated methods that have been explored in the general econometrics literature (Baltagi, 2008). In particular, we choose the method proposed by Levin *et al.* (2002), which considers unit root tests in small panels[5] like ours. If nonstationarity is present, we then use the Westerlund (2007) test for the null of no cointegration. If this null is rejected, we then estimate the cointegrated panel after choosing between pooled mean-group (PMG) and mean-group (MG) estimators using the Hausman test. We do find stationarity (consistent with earlier studies (Narayan and Popp, 2012)) so that there is no need to test cointegration.

3.3 Performance Criteria for Economic Forecasting

A fundamental challenge is how to measure the performance of a particular projection model into the future, rather than simply the “fit” of an estimation model into the past data. Three formal criteria are well established in the field of economic forecasting — accuracy, precision, and certainty (Clements and Hendry, 2004, 2011) — among

[5]Other methods such as Im *et al.* (2003); Harris and Tzavalis (1999) require both (infinitely) large N and T for the test statistic to have a well-defined asymptotic distribution.

which, the first is usually treated as the overriding criterion. Accuracy refers to the unbiasedness of the forecast Y^f[6]. Assuming that the intercept equals 0 for simplicity, accuracy is determined by the unbiasedness of estimated parameters: the estimated coefficient of X (i.e., b). Precision means the small size of the variance (σ^2) of the forecast Y^f. Accuracy and precision are often combined into one indicator. In general, there are three alternative groups of such indicators: mean error (ME), mean absolute error (MAE), and mean square error (MSE) (Makridakis *et al.*, 1998)[7]; that is,

$$ME = \sum_{t=T+1}^{T+F} \frac{(Y_t - Y_t^f)}{F},$$

$$MAE = \sum_{t=T+1}^{T+F} \frac{|Y_t - Y_t^f|}{F},$$

$$MSE = \sum_{t=T+1}^{T+F} \frac{(Y_t - Y_t^f)^2}{F}. \tag{3.2}$$

The sizes of these indicators, unfortunately, depend on the scale of the data. For example, if one used them to select better-performing models where the data are mixed with unlogged and logged scales, then the logged $(Y_t - Y_t^f)$ would be no doubt much smaller than the unlogged $(Y_t - Y_t^f)$ in their numerical values, and one would always mistakenly choose the model using logged data regardless of the model specifications. Because of this, this group of indicators cannot facilitate comparisons across models with different scales of expenditure, which is one of our interests in this

[6] "The notion of unbiasedness, whereby forecasts are centered on outcomes, is used in technical analyses to measure accuracy; whereas that of small variance, so only a narrow range of outcomes is compatible with the forecast statement, measure precision ... When (squared) bias and variance are combined one-for-one, we obtain the commonly-reported mean square forecast error (MSE)." (Clements and Hendry, 2004) (p.5)

[7] F refers to the number of forecast periods.

paper.

An alternative group of indicators can measure the relative size of forecast error, and hence enable comparisons among various scales of data: percentage error (PE), mean percentage error (MPE), mean absolute percentage error (MAPE), and mean squared percentage error (MSPE); that is,

$$PE = 100 \times \frac{(Y_t - Y_t^f)}{Y_t},$$

$$MPE = \sum_{t=T+1}^{T+F} \frac{(PE)}{F},$$

$$MAPE = \sum_{t=T+1}^{T+F} \frac{|PE|}{F},$$

$$MSPE = \sum_{t=T+1}^{T+F} \frac{(PE)^2}{F}. \tag{3.3}$$

This group of indicators is further subject to two weaknesses. First, if the scale of data includes values close or equal to zero, then the denominator in PE is close to zero, which leads PE to approach infinity. The second disadvantage — shared by both groups of indicators above — is that they represent pure numbers that, without a reference (e.g., naïve model), cannot indicate relative gains in performance using a particular model. For example, ME just gives a mean number of $(Y_t - Y_t^f)$ and MPE offers that of $\frac{(Y_t - Y_t^f)}{Y_t}$. Both measure the distance the forecast Y_t^f of a formal model from the actual reality Y_t rather than from that of another formal model for the very purpose of selecting the better-performing model. The conventional way to proceed is to separately calculate the corresponding MAE, as done by Getzen and Poullier (1992), of the naïve model which uses the most recent observation available Y_t as a one-step forecast Y^f_{t+1}, then compare the indicator based on the naïve model against a formal model, and then select the better performing one. When candidate models are many, this procedure becomes lengthy and inconvenient. Theil′s U statistic, in contrast, is a formal and convenient indicator that builds in the reference naïve method (Theil, 1966; Makridakis *et al.*, 1998) and hence can directly compare the relative performance of candidate models all at once. It also has other favorable characteristics — it penalizes large errors and acts as a compromise between the absolute and relative measures; that is,

$$U = \sqrt{\frac{\sum_{t=T+1}^{T+F}(FPE_{t+1} - APE_{t+1})^2}{\sum_{t=T+1}^{T+F}(APE_{t+1})^2}}, \tag{3.4}$$

where Forecast Percentage Error: $FPE_{t+1} = \frac{Y^f_{t+1} - Y_t}{Y_t}$ and Actual Percentage Error: $APE_{t+1} = \frac{Y_{t+1} - Y_t}{Y_t}$. Hence the Theil′s U formula can be simplified as

$$U = \sqrt{\frac{\sum_{t=T+1}^{T+F}\left(\frac{Y^f_{t+1} - Y_{t+1}}{Y_t}\right)^2}{\sum_{t=T+1}^{T+F}\left(\frac{Y_t - Y_{t+1}}{Y_t}\right)^2}}. \tag{3.5}$$

Similar to PE, the FPE and APE[8] respectively measure the predicted relative change of a formal model and the actual relative change of the naïve method. Resembling MSE, both FPE and APE are squared and hence more sensitive to large errors. Finally, the reference of the naïve method is successfully built into APE acting as the denominator of U. Therefore, Theil′s U statistic integrates the methodological advantages of the two alternative groups of indicators, though it does not discriminate between signs like MSE (Theil, 1966). The range and magnitude of U values

[8]Actual refers to the naïve method which uses the actual past value, Y_t, directly for 1-step forecast value, Y^f_{t+1}.

provide the judgment criteria for forecast accuracy and precision of a model, relative to the naïve reference. If U $>$ 1, the naïve method is preferred to a formal forecast model; if U $=$ 1, both methods are equally good; if U $<$ 1, a formal model is preferred over the naïve method. The closer the U-value is to zero, the more accurate and precise the formal model is. Some may argue that this statistic posits a very simple reference so that it is not hard for a model to perform better than the reference. But this argument is not a concern in our context since what we are comparing are the U statistics between two or more forecast models given the common reference.

Finally, there is an alternative measure of accuracy and precision that was developed by Diebold and Kilian (2001). They noticed that some variables could be intrinsically more difficult to predict than others. A large gap between forecasted and actual values of such a variable is mainly because the future of the series depends less on its past, rather than the forecaster's failure. It proposes the alternative measure of accuracy and precision as a ratio of the error of a short-run forecast to that of a long-run forecast. As the authors illustrate, this measure is very similar to that of Theil′s U statistic. The main difference is the time horizon. Theil′s U evaluates one-step forecast accuracy of a forecast model relative to that of the naïve model, whereas this alternative measure assesses the one-step forecast accuracy of a forecast model relative to its $k-$step accuracy. So the major strength of this alternative measure is that it allows for different forecast horizons. However, the health expenditures THE and PHE are much less volatile variables, and the international data are almost yearly. That is, HEs are not less-intrinsically-predictable variables and hence are not suitable for applying this alternative measure of accuracy. Therefore, we use Theil′s U statistic as the key performance criterion throughout this paper.

Uncertainty refers to the randomness of the actual Y in the future, which may (i.e., known/measurable uncertainty) or may not (i.e., unknown/immeasurable uncertainty) be captured by the confidence interval of the forecast Y (Clements and Hendry, 2004, 2011). Unknown uncertainty, as noted in the introduction, is due to the limitation of human cognition, and hence cannot be incorporated into models of estimation and prediction. For example, the probability of the Earth being attacked by aliens is beyond human knowledge. Note that this unknown uncertainty is not pertinent to the problem of intrinsic difficulty to predict as described in the preceding paragraph, which refers to the wider confidence interval of volatile variables and is the known uncertainty. The known uncertainty is often measured as the confidence interval (i.e., the range of certainty)[9] conditional on type I error (α) and on the i.i.d. assumption; that is,

$$(Y^f - z_{\alpha/2}\sigma, \quad Y^f + z_{\alpha/2}\sigma), \tag{3.6}$$

where σ is replaced by MSE that is obtained from the sample data. It means that the actual Y in the future would have $(100\% - \alpha)$ probability falling within this forecast interval. It remains valid particularly for more than 1-step forecast only when the i.i.d. assumption holds.

3.4 Strategy of Analysis

We employ an analytical strategy consisting of five stages. In Stage 1 we diagnose the properties of each single time series of THE and PHE containing more than 30

[9] Alternative measurement could be the anticipated probability distribution of a volatile outcome e.g., inflation visualized like the well-known "fan chart" (Bank of England, 2015).

observations between 1960 to 2008 for OECD member countries. Each series is separately measured at two levels — per capita and total; and in three currency units — current national currency units (NCU, nominal), NCU adjusted by GDP deflator in 2000 (NCU, real), and, for the purpose of international comparison, adjusted by Purchasing Power Parity (International Dollars, IntD, PPP). If the series was non-stationary in mean or variance, we transformed it into stationary by differencing or taking the natural logarithm.

In Stage 2, following the rule of thumb in the field of economic forecasting, we estimate parameters in proposed models from the first two-thirds (e.g., 1960-1993) of the historical data (called initialization set). We then use these estimates to predict the dependent variable in the later one-third of data (called test set, e.g.,1994-2008).

In Stage 3, around these within-sample forecasts from various methods, we calculate Theil′s U values and forecast intervals to compare and contrast their accuracy, precision, and certainty. We then select the better-performing models (in terms of both how expenditures are measured and the projection models) based on such comparisons.

In Stage 4 we apply the selected models to the full set of data, to re-estimate parameters, and based on these estimates, conduct out-of-sample forecasts for 2015–2025. If such forecasts were logarithmized, then we transform them back to their original scale using and comparing two alternative ways of exponentiation[10]. One way is to directly exponentiate the logged estimated mean of expenditure y for the unlogged mean of forecast e^y; and the logged estimated standard error of the forecast σ for the unlogged standard error of the forecast $\sqrt{e^{\sigma^2}}$. The alternative way

[10]To get a sense that how scaling down (i.e., logged) of the data (i.e., expenditures) can distort the forecast of the scaled-up data (i.e., exponentiate the logged expenditures).

is to exponentiate both the logged estimated mean and the logged variance for the unlogged mean of forecast $e^{y+(\sigma^2)/2}$, and for the unlogged standard error of forecast $\sqrt{e^{2y+\sigma^2}(e^{\sigma^2}-1)}$ (Lin, 2012).

To further reduce the, randomness of individual forecasts obtained from each preferred model, in Stage 5, we combine the out-of-sample forecasts across models. The literature shows that either simple arithmetic average or complicated combination (different weights) of forecasts considerably increase accuracy and precision (Clemen, 1989; Elliott and Timmermann, 2008), and reduce uncertainty (Makridakis and Winkler, 1983; Makridakis *et al.*, 1998), compared to one individual forecast. The most commonly used combination is to take the average.

3.5 Data

All health expenditure and macroeconomic statistics are from the OECD System of Health Accounts (SHA) database. The data are collected annually for 34 member countries. Although the length of the period varies slightly among countries, the 1960-2008 time series is much longer than any other international databases available. Therefore, we apply SHA to both time series and panel data models. For the completeness of available data for most countries, we choose types of health expenditures mainly classified by financing agents, and focus specifically on THE and PHE (classified by financing sources). These two types of expenditures are of considerable policy interest (OECD, 2011). We include countries and years that together can provide the largest number of observations and create strongly balanced panels. Specifically, we choose 20 countries from 1972 to 2008 (T=37) for THE and thus have 740 observations. The sample includes 18 countries over 1980-2008 (T=29) to provide

522 observations for PHE (Table 3.1). Three alternative datasets are also identified: World Health Organization (WHO) National Health Accounts (NHA) from 1995 to 2009; World Bank (WB) World Development Index (WDI) data for 1970–2009; International Monetary Fund (IMF) Government Finance Statistics (GFS) incompletely covering 2002–2008. These series, unfortunately, are too short to provide stable parameter estimates by time series approaches, we only use them to assess the quality of data from OECD-SHA. All estimation was performed using STATA 12.0.

3.6 Results

3.6.1 Comparison of Within-Sample Forecast Accuracy and Precision

In the interest of international comparison, $ refers to the International Dollar, which is the current national currency unit adjusted by PPP. We choose countries (N) and years (T) that together provide the largest number of observations and strongly balanced panels. So the N and T differ for the panels of THE and PHE, and the expenditure amounts cannot be compared between the two panels. This is because OECD countries differ largely in the total population. Table 3.1 reports the descriptive statistics of variables of 20 countries (N=20) during 1972–2008 (T=37) for forecasting THE. The mean of THE is $12,430 million in total and $1,210 per capita. The mean of GDP is $66,043 million in total and $15,975 per capita. The mean of the total population is 10,392 in thousands. The female and the elderly respectively account for about 50.87% and 13.22% of the population on average for these OECD countries.

Table 3.1 also reports the descriptive statistics of variables of 18 countries (N=18) during 1980–2008 (T=29) for forecasting PHE. The mean of PHE is \$12,539 million in total and \$1,077 per capita. The mean of GGR is \$133,119 million in total, and \$9,119 per capita. The mean of the total population is 11,610 in thousands. The female and the elderly respectively account for about 50.80% and 13.01% of the population on average for these OECD countries.

Tables 3.2 and 3.3 provide the indicators of forecast performance of different measurements of Canadian THE and PHE as well as the performance of statistical smoothing models versus econometric time series models. We only conduct comparisons of expenditure measurements using Canadian series data since each country has one time series, and here we are interested in how (time) series models perform differently rather than how different countries' data perform differently. The fact that all Theil′s U values for statistical and time series data models are smaller than 1 indicates that these models are all better predictors than the naïve model. We report the results of four comparisons: (a) total vs. per capita measures, (b) logged vs. unlogged scales, (c) alternative currency units of expenditures, and (d) one statistical smoothing and three time series models.

First, for all time series models except VAR in only one case, total measures overall produce smaller Theil′s U values than per capita measures. For example, in the last column of Table 3.2, the Theil′s U value 0.0043 for totals in the first row of the first panel is slightly smaller than 0.0064 for per capita in the first row of the third panel, conditional on logged scale and international dollars produced by the VAR model.

Second, the "logged" scale provides smaller Theil′s U values than the "unlogged" scale. For instance, comparing the first rows of the first and third panels with those of the second and fourth panels, the Theil′s U values 0.0043 and 0.0064 for logged totals and per capita respectively are smaller than 0.1502 and 0.1311 for the unlogged counterparts, in international dollars produced by the VAR model. Therefore, we prefer “logged total” for subsequent out-of-sample forecasts.

Third, other things equal, the choice among three currency units differs across models. International dollars at current PPP usually outperforms in the VAR model. For example, in the first row of Table 3.2, comparing the last column with the ninth and eleventh columns, the U value 0.0043 produced by VAR is smaller than U values 0.0072 and 0.0204 respectively by ARIMA-dynamic and double exponential (DE) models for logged totals. Whereas NCU adjusted by GDP deator in the middle often produces smaller U values in other time series models than VAR (e.g., 0.0069 provided by VAR is larger than 0.0009 and 0.0015 respectively by ARIMA-static and dynamic models for logged totals). Both NCU adjusted by PPP and by GDP deflator are better than NCU in nominal terms. We choose international dollars for out-of-sample forecast throughout this paper to facilitate international comparison using OECD data and because projections of inflation provided by financial forecasters for the out-of-sample forecast may later introduce extra error to forecasts using GDP-deflator.

Fourth, comparing three groups of performance indicators, Table 3.2 shows that, as expected, both U and the relative measures — MAPE and RMSPE — are less sensitive to the scale of PHE series than the absolute measures. Although the numerical values of certain indicators change substantially across models, the order of model performance based on each of these indicators, conditional on data scale and currency unit, is stable. That is, ARIMA-static (AS) $<$[11] VAR $<$ ARIMA-dynamic (AD) $<$

[11] $<$ represents less than in terms of the numerical value of U-statistic produced by models.

double exponential smoothing (DE) for logged scale, both totals and per capita, in IntD. Table 3.3 shows similar results for THE except that VAR seems on par with double exponential smoothing. That is, ARIMA-static $<$ DE$<$ VAR $<$ ARIMA-dynamic for logged totals; and ARIMA-static $<$ VAR $<$ DE $<$ ARIMA-dynamic for logged per capita.

Tables 3.4 and 3.5 report results for THE and PHE of three further comparisons: (a) logged total vs. logged per capita, (b) logged vs. exponentiated in two ways, and (c) time series vs. panel data models using international dollars[12]. The first panel shows that the Theil′s U values for all models are smaller than 1, which suggests all are better than the naïve model. Comparing the first and third rows of each panel reporting Theil′s U, MAPE, and RMSPE values, we find that logged totals again have smaller values than logged per capita. The static panel data model with FE estimator (SP-FE) seems especially sensitive to the two alternative ways of exponentiation. For example, comparing the first rows of the fifth and eighth numerical columns in Table 3.4, the U value of the static panel-FE model for THE increases from 0.0051 to 0.0442 under the strong assumption that the standard error of forecast equals zero, but explodes to 0.9336 when this strict assumption is relaxed. Because 0.9336 approaches 1, the static panel-FE model becomes the worst among projection models and not much better than the naïve model. In comparison, smoothing and time series models in the first through fourth numerical columns overall outperform panel data models for both logged and exponentiated expenditures.

Finally, comparing the first through seventh numerical columns in Table 3.4, for

[12]Because soon we will provide and rank out-of-sample forecasts of OECD countries, the within-sample comparisons of measures and models chosen based on Canadian series should use the currency unit that allows international comparison.

THE exponentiated without the assumption of zero estimated error[13], the order of performance based on Theil′s U follows as ARIMA-static < DE < VAR < Dynamic panel-xtabond[14] estimator = Dynamic panel-xtdpd[15] estimator < ARIMA-dynamic < Static panel-FE estimator. Table 3.5 shows that PHE has similar results. However, the double-exponential smoothing model performs worse, which may suggest its sensitivity to the length of initialization set as the first two-thirds of past data[16]. What's more, although the three estimators for the dynamic panel model all passed specification tests, only the Arellano and Bover/Blundell and Bond estimators outperform time series models.

3.6.2 Combined Out-of-Sample Point Forecasts, and Forecast Intervals

We report country-specific expenditures (Table 3.6) and ranks (Table 3.7) of THE and PHE in international dollars per capita to facilitate cross-country comparison. Table 3.6 reports both the point forecast and the upper and lower bounds[17] of the forecast for each country in 2015, 2020, and 2025. To facilitate comparison, Table 3.7 presents the results ranked by point forecast expenditures.

To compare the point forecasts of most interest, Table 3.7 shows that in 2015 the US is forecast to spend the largest amount of THE per capita — up to $10,413 — about 5–6 times greater than the point forecast for Turkey. Following the US,

[13]To get a sense that how scaling down (i.e., logged) of the data (i.e., expenditures) can distort the forecast of the scaled-up data (i.e., exponentiate the logged expenditures).
[14]xtabond represents Arellano and Bond estimator.
[15]xtdpd refers an alternative estimator to Arellano and Bond, as well as Arellano and Bover/Blundell and Bond system estimators.
[16]This is the tradition of the forecasting literature.
[17]Defined in the confidence interval introduced in section 3.

European countries such as Norway (NOR), the Netherlands (NLD), and Germany (DEU) consistently rank in the top 5–10 throughout the forecast horizon, with a level of \$5,000–9,000. Among them, the Netherlands climbs much faster than others, replaces the US at \$15,816 in 2020, and stays on top at \$28,488 through to 2025. Others, such as Iceland (ISL), conversely, are forecast to decrease THE down to the bottom of the rank in the medium run. English-speaking countries except the US consistently locate in the middle. Specifically, Canada is forecast to increase THE to \$6,101 in 2015, and up to \$8,347 and \$11,444 in 2020 and 2025 respectively, whereas Australia spends slightly higher (at \$4,884) than GBR (at \$4,857) in 2015, but lower (at \$8,148) than GBR (at \$9,036) in 2025.

PHE per capita forecast differs from THE. Norway, followed by the US, is forecast to become the largest public financier of health care at \$6,662 and \$12,835 in 2015 and 2025. Whereas the US' PHE is forecast to reach at \$11,046 in 2025. The Netherlands' PHE, similar to its THE, rises much faster again than that of other countries up to \$10,131 in 2025 and ranks third highest in absolute amount. Other English-speaking countries again remain in the middle. Canada's and Britain's PHE both rise slightly toward \$6,000 and \$8,500 in 2020 and 2025, higher than Australia's PHE of about \$4,600 and \$6,200 for the same years.

The ratio of PHE to THE, however, shows a different picture from the absolute amounts noted above. In the rank of per capita PHE as the proportion of per capita THE, GBR is forecast to have the persistently largest public share (i.e., near 90%) of THE until 2020. European countries consistently rank in the top 5-10, while the Netherlands is forecast to have the second lowest public share[18]. Meanwhile, the US'

[18]About 51% of its THE in 2015, and quickly shrinks public support down to 43% in 2020 and 36%, lower than the US', in 2025.

PHE accounts for half of its THE and incrementally grows to 53%, 59%, and 65% of THE in 2015, 2020, and 2025. Canada's PHE accounts for a large share (i.e., 80%)[19] and remains slightly lower in the future around 74%. Japan's PHE share is expected to climb from 84% in 2015 up to 92% in 2025. Whereas Portugal's drops from 53% in 2015 to 34% in 2025 toward the bottom of the rank.

Figures 3.1–3.4 show that forecasts in total differ slightly from those in per capita above. It is the two large economies of Japan and Germany, rather than the Netherlands, who are expected to follow the US in spending the most on health care. Such patterns are also observed in PHE forecasts in absolute total terms. The US, Canada, and European countries such as Iceland show apparent reductions in THE (Figures 3.1 and 3.2) and PHE (Figures 3.3 and 3.4) in 2009 possibly because their 2008 data already reflected an economic slow-down. Whereas Australia's THE appears relatively smooth and shows a slightly increasing curve whose positive slope is smaller than that of Japan. South Korea, as one of the emerging Asian economies, possesses the steepest upward curve. Iceland's PHE, in particular, is flatter than its THE curve, whereas the Britain's and the US' PHEs show steeper downward curves in 2009 than their THEs. These observations coincide with the stringent budget constraints of these countries since the middle of 2008.

[19] This is higher than Canadian historical records at an average 70% during 2000–2012. Recall this is the forecast combined from four smoothing and time series models. The individual forecast of PHE/THE provided by individual model DE, HW, AD, and VAR respectively is 74%, 97%, 71%, and 80%. Obviously, the final combined 80% comes from the VAR and especially the HW model, which corresponds to HW's high sensitivity to the end points of data used to forecast.

3.6.3 Data Discrepancy

Working with these international databases has identified a number of potential quality issues. Below we illustrate some particularly large[20] and systematic discrepancies in comparison of the same data across different databases. Table 3.8 shows that PHE provided by the OECD and WHO databases for Switzerland are both 24 billion NCU higher than that provided by the IMF database in 2008, by a factor of 3. Similarly, OECD and WHO database derived PHEs for Denmark are respectively 16 and 7 NCU billion higher, on average, over 2002–2009. Out-of-pocket payment (OOP) provided by the OECD database for Japan, compared to that provided by the WHO database, is on average 73 billion NCU lower, ranging from 671 billion lower in 1998 to 852 billion higher in 2006. GDP series for South Korea provided by the OECD and UN databases are consistent, but both are 14,956 billion NCU higher than that provided by the WB database. WHO data for this series is even higher (i.e., 19,976 billion) on average over 1995–1999, but all four databases for this series suddenly harmonize beginning in 2000. The PHE as a percentage of GGE for Switzerland offered by the OECD or WB databases is about 14 percentage points higher than that provided by the IMF database for 2007. Such a percentage offered by the OECD or WB databases, compared to that for the WHO database, is, on average, 10 percentage points higher for Chile and lower for Finland. On the other hand, OECD, WHO, WB data only harmonize for South Korea after the year 2000, for Chile after 2003, and for Denmark before 2003. This suggests possible sources of data discrepancy including data management changes and exogenous events occurring in these countries during these years.

[20]Large discrepancy in international databases means the correlation coefficient between two amounts of the same variable in two databases is below 0.5.

3.7 Discussion and Conclusion

Our results indicate that, contrary to Getzen and Poullier (1992), complicated econometric (esp., static) panel data models perform worse than simpler statistical (e.g., smoothing) and econometric time series models for forecasting. This finding is inconsistent with the early findings by Getzen and Poullier (1992) and responds to their call for forecasting based on more years (nearly 20, in our case) of data. This finding is, however, consistent with empirical findings from the broader literature on economic forecasting (Armstrong, 1978; McNees, 1986). Therefore, to forecast, more complicated econometric models — except VAR — generally do not do better than simpler ones such as exponential smoothing. This implies that a (usually complex) model that better fits historical data does not guarantee accurate and precise post-sample forecasts. This chapter also finds that some methods perform more accurately for short horizons (e.g., exponential smoothing) while others are more appropriate for medium horizons (e.g., ARIMA, VAR). This is also consistent with studies from the broader literature on forecasts (Fildes and Makridakis, 1995; Fildes *et al.*, 1998).

The recent literature review (Astolfi *et al.*, 2012) on health expenditure forecasting suggests that with better computing power and more refined data, the future of forecasting is complicated micro models. But micro panel data models can forecast worse than time series and smoothing models. Our paper confirms this. Astolfi *et al.* (2012) also suggest structural models (e.g., CGE) for forecasting that specify explicit causal-effect mechanisms among variables. In comparison, the models used in this paper are reduced-form and hence "atheoretical", which is fine since our objective is to obtain accurate and precise forecasts of future spending rather than to understand past determinants. It is important to carefully distinguish the objective of simply getting accurate forecasts of future spending from modeling from that of understanding determinants of expenditure growth. Arriving at such an understanding of the determinants of expenditure growth requires considerably more data and effort, and may still do worse with pure forecasting. Our paper confirms this. Thus, different kinds of projection models may work better in different situations and respond to different demands from policy makers. For instance, theoretical models are more suitable for "if ... then" policy questions. For health policy makers in the Ministry of Health, forecasts based on a health-sector-only structural model would be enough. Whereas for those who are responsible for more general issues (e.g., analysts in Ministry of Finance) across all major sectors (not only health but also education, manufacturing, etc.), a CGE model is demanded. Unfortunately, a well-agreed theoretical model for health expenditure at the macro level has not yet been well established, though there are prototypes of CGE models (Astolfi *et al.*, 2012). This might be why articles on health expenditure forecasts are few, although cost estimates are many. This inquiry urges, as Gerdtham and Jonsson (2000) called for more than a decade ago, both macroeconomic theories and macro structural models for health expenditure (Heckman, 2000; Carnot *et al.*, 2011).

Our paper also contributes to the call made by Astolfi *et al.* (2012) for more rigorous methods, for more transparency, and for better assessment of performance. Policy makers could choose time series models that treat the system as a "black box" and do not attempt to discover the factors affecting its behavior. Univariate time series methods also suit when the main concern is to forecast what will happen to expenditures, not why or how it happens. If the latter is important, explanatory forecasts based on multivariate models from time series data such as VAR or panel data will be preferred. Elliott and Timmermann (2008), however, emphasize that time series models can be unstable so that one cannot rely on the same dominant model in different historical samples. Hendry and Hubrich (2012) similarly conclude that after introducing the uncertainty of a variable such as the Consumer Price Index (CPI), aggregate forecast using aggregate data performs less accurately than that using summarized disaggregate information. However, these are not concerns in terms of our objective — to forecast health expenditures, which are less volatile and hence less uncertain than variables like stock price and CPI. Nonetheless, we recommend that health expenditure forecasters who have chosen one kind of time series model continually compare its performance with that of other kinds of time series models. Although we argue Theil′s U statistic is the superior measure of the performance of projection models, we are aware that judgment of models might differ when using alternative indicators (Makridakis *et al.*, 1993). Characteristics such as randomness and frequency of data matter for forecasting as well. For annual HE data like ours in which the trend dominates cyclical fluctuations, and little randomness is present, we recommend Holt-Winters non-seasonal and its alternative — double exponential smoothing, and time series models. Finally, we are aware that the international databases have been updated and regularly revised so that particular patterns of the data might be susceptible to such revisions and lead to models' misspecification. We recommend using the most recently updated data.

Budgeting and strategic planning for the short and medium run, and fiscal sustainability in the long run can only be achieved with sound forecast modeling, monitoring, and modifying (Makridakis *et al.*, 1998). Forecasts provided by either the academia or the policy arena need to be updated for periodic changes and tracked on records of performance for the purposes of modifying and improving methods in both theory and practice.[21] Health ministers usually favor higher growths of the public health budget, but finance ministers prefer lower. No matter how objectively health expenditure forecasting is improved, decisions concerning allocation of monetary resources to the health sector inevitably involve the human judgment that may be inuenced by bargaining power and political considerations hence have little to do with objectivity.

[21]Recall that it took decades for Samuelson to withdraw his previous prediction from his textbook that Soviet GNP would exceed that of the United States by as early as 1984 or perhaps by as late as 1997 (Samuelson, 1980). "The future will prove that anything could be wrong" reflected by Thomas Getzen.

Table 3.1: Variable statistical summary of panel data for OECD countries

Total health expenditure (THE), 20 countries, 1972-2008				Public health expenditure (PHE), 18 countries, 1980-2008			
Variable	Observations	Mean	Std. Dev.	Variable	Observations	Mean	Std. Dev.
THE_{MIntD}	739	12430.40	6.68	PHE_{MIntD}	521	12539.01	6.13
THE_{PCIntD}	716	1210.03	2.30	PHE_{PCIntD}	506	1076.75	2.23
GDP_{MIntD}	740	66042.66	5.85	GGR_{MIntD}	392	133119.17	4.89
GDP_{PCIntD}	740	15975.32	1.94	GGR_{PCIntD}	392	9118.86	1.88
POPTOT	740	10392.09	5.25	POPTOT	522	11609.74	5.88
POPFE	740	50.87	1.01	POPFE	522	50.80	1.01
POPELDER	740	13.22	1.21	POPELDER	522	13.01	1.30

Notes: According to definitions from OECD et al. (2011), THE refers to total spending executed on health goods and services whose primary purposes include maintenance, restoration or enhancement of health (PHE). GGR is general revenues of all levels of government. POPTOT refers to total population size, in thousands. POPFE is the proportion of the female population. POPELDER is the proportion of the elderly aged at 65 and above.
Monetary variables are measured in millions of International Dollars, IntD, in total and per capita respectively. MIntD refers to million international dollars (IntD). PCIntD means per capita IntD.
The observations of THE and PHE differ, because we include countries and years that together can provide the largest number of observations and create strongly balanced panels.

Table 3.2: Comparison of forecast performance of measurements of health expenditures using alternative indicators, Public health expenditure (PHE), Canada, within test set 1997–2008

		NCU, Current Price				NCU, GDP deflator				IntD, Current PPP			
Measure of series	Performance indicator	DE	AS	AD	VAR	DE	AS	AD	VAR	DE	AS	AD	VAR
Total, logged	Theil's U	0.0212	0.0019	0.0108	0.0192	0.0176	0.0009	0.0015	0.0069	0.0204	0.0012	0.0072	0.0043
	MAE	0.4083	0.0369	0.2322	0.3423	0.3453	0.0132	0.0306	0.1255	0.3873	0.0226	0.1507	0.0739
	MSE	0.2209	0.0019	0.0655	0.1807	0.1514	0.0004	0.0012	0.0235	0.1971	0.0007	0.0275	0.0090
	MAPE	3.5765	0.3289	2.0384	2.9930	3.0543	0.1185	0.2717	1.1089	3.4534	0.2055	1.3464	0.6588
	RMSPE	4.0928	0.3857	2.2330	3.6950	3.4298	0.1718	0.3019	1.3505	3.9342	0.2339	1.4740	0.8386
Total, unlogged	Theil's U	0.2528	0.0094	0.2645	0.1322	0.2295	0.0098	0.2783	0.1306	0.2403	0.0140	0.2481	0.1502
	MAE	28925	1377.3	29955	161501	25562	1128.5	29648	15969	22816	1871.3	23392	15081
	MSE	1.2e+09	2490985	1.3e+09	3.9e+08	8.5e+08	2272782	1.2e+09	3.2e+08	7.4e+08	3734240	7.8e+08	3.3e+08
	MAPE	31.698	1.9085	32.787	17.500	31.485	1.7140	36.335	19.821	30.542	2.9058	31.285	20.080
	RMSPE	35.449	2.3236	36.742	19.935	34.656	2.3974	40.519	21.423	34.028	3.0232	34.910	22.567
Per capita, logged	Theil's U	0.0313	0.0023	0.0138	0.0070	0.0255	0.0012	0.0031	0.0082	0.0301	0.0017	0.0094	0.0064
	MAE	0.4145	0.0294	0.2090	0.0799	0.3445	0.0124	0.0453	0.0997	0.3917	0.0221	0.1377	0.0747
	MSE	0.2272	0.0013	0.0529	0.0117	0.1505	0.0004	0.0024	0.0160	0.2013	0.0006	0.0230	0.0095
	MAPE	5.2156	0.3797	2.6376	1.0019	4.3949	0.1609	0.5799	1.2689	5.0571	0.2923	1.7829	0.9620
	RMSPE	5.9548	0.4622	2.8804	1.3449	4.9320	0.2481	0.6264	1.6054	5.7479	0.3371	1.9475	1.2419
Per capita, unlogged	Theil's U	0.2516	0.0093	0.2734	0.1041	0.2338	0.0081	0.2768	0.1199	0.2382	0.0144	0.2591	0.1311
	MAE	912.82	42.177	973.78	413.52	824.82	28.891	939.30	473.10	718.11	59.427	766.76	428.84
	MSE	1167165	2382.6	1341489	250206	876669	1536.7	1167010	271973	718434	3853.2	827665	255422
	MAPE	32.215	1.8138	34.291	14.493	32.576	1.2822	36.969	18.852	30.955	2.9277	32.976	18.513
	RMSPE	35.991	2.2243	38.474	16.390	35.917	1.2822	41.246	20.242	34.447	3.0860	36.866	20.514

Notes: NCU refers to national currency unit. DE refers to Double-Exponential smoothing model; AS refers to ARIMA model with static forecast; AD refers to ARIMA model with dynamic forecast. VAR refers to VAR model. RMSPE means the square root of MSPE.

Table 3.3: Comparison of forecast performance of measurements of health expenditures using alternative indicators, Total health expenditure (THE), Canada, within test set 1997–2008

		NCU, Current Price				NCU, GDP deflator				IntD, Current PPP			
Measure of series	Performance indicator	DE	AS	AD	VAR	DE	AS	AD	VAR	DE	AS	AD	VAR
Total, logged	Theil′s U	0.0108	0.0013	0.0209	0.0128	0.0167	0.0006	0.0024	0.0106	0.0050	0.0014	0.0180	0.0131
	MAE	0.1905	0.0240	0.4577	0.2658	0.3400	0.0105	0.0496	0.2246	0.0557	0.0266	0.3856	0.2710
	MSE	0.0618	0.0009	0.2430	0.0901	0.1446	0.0002	0.0030	0.0592	0.0101	0.0010	0.1720	0.0910
	MAPE	1.6175	0.2092	3.9280	2.2742	2.9164	0.0911	0.4268	1.9293	0.5131	0.2353	3.3665	2.3606
	RMSPE	2.0981	0.2610	4.2051	2.5489	3.2519	0.1234	0.4695	2.0843	0.8750	0.2839	3.5995	2.6096
Total, unlogged	Theil′s U	0.1644	0.0075	0.1387	0.1240	0.2067	0.0129	0.0963	0.1578	0.1406	0.0119	0.1268	0.1085
	MAE	23595	1201.1	20290	18072	33491	2528.4	16911	26986	16773	1800.0	15264	12952
	MSE	1.1e+09	2819968	7.8e+08	6.4e+08	1.5e+09	7789656	3.8e+08	9.1e+08	5.4e+08	4803428	4.5e+08	3.4e+08
	MAPE	17.903	1.1535	15.417	13.687	28.971	2.3897	14.586	23.514	15.516	2.0579	14.132	11.950
	RMSPE	22.944	1.7091	19.701	17.665	31.773	2.6855	16.147	25.361	19.885	2.4718	18.089	15.580
Per capita, logged	Theil′s U	0.0143	0.0016	0.0261	0.0035	0.0240	0.0008	0.0043	0.0103	0.0100	0.0018	0.0226	0.0024
	MAE	0.1761	0.0188	0.4062	0.0459	0.3411	0.0099	0.0635	0.1424	0.1204	0.0246	0.3430	0.0318
	MSE	0.0525	0.0007	0.1889	0.0032	0.1454	0.0002	0.0048	0.0274	0.0246	0.0008	0.1343	0.0014
	MAPE	2.1205	0.2355	4.9572	0.5635	4.1646	0.1219	0.7767	1.7387	1.4870	0.3127	4.2899	0.4022
	RMSPE	2.7381	0.3244	5.2680	0.6858	4.6395	0.1698	0.8475	2.0124	1.9166	0.3699	4.5529	0.4820
Per capita, unlogged	Theil′s U	0.1539	0.0078	0.1350	0.0998	0.2097	0.0073	0.0765	0.1369	0.1277	0.0110	0.0931	0.0894
	MAE	713.09	37.802	635.27	476.49	1077.0	40.151	436.87	759.895	493.17	50.430	369.568	348.04
	MSE	934016	2990.9	738657	428893	1480130	2520.1	243044	700197	451902	4072.3	252996	236671
	MAPE	17.472	1.1164	15.590	11.695	29.870	1.2281	12.115	21.241	14.743	1.8107	11.096	10.404
	RMSPE	22.309	1.6775	19.836	14.947	32.790	1.5877	13.338	22.793	18.802	2.2476	14.053	13.468

Notes: NCU refers to national currency unit. DE refers to Double-Exponential smoothing model; AS refers to ARIMA model with static forecast; AD refers to ARIMA model with dynamic forecast. VAR refers to VAR model.

Table 3.4: Comparison of forecast performance of time series models versus panel data models using alternative indicators, THE, logged total and per capita expenditures adjusted by PPP, Canada, within test set 1997–2008

	Series Measure	DE	AS	AD	VAR	SP-FE-exp1	DP1-exp1	DP3-exp1	SP-FE-exp2	DP1-exp2	DP3-exp2
Theil′s U	total logged	0.0050	0.0014	0.0180	0.0131	0.0051	0.0149	0.0149	0.0051	0.0149	0.0149
	total exponentiated	0.1071	0.0142	0.2423	0.1831	0.0442	0.2055	0.2055	0.9336	0.2333	0.2333
	per cap logged	0.0100	0.0018	0.0226	0.0024	0.0072	0.0174	0.0174	0.0072	0.0174	0.0174
	per cap exponentiated	0.0925	0.0129	0.2092	0.0190	0.0632	0.1425	0.1425	0.9470	0.1583	0.1583
MAE	total logged	0.0557	0.02657	0.3856	0.2710	0.0839	0.2590	0.2590	0.0839	0.2590	0.2590
	total exponentiated	13086	2288.1	48404	32724	11860	60483	60483	3847245	60545	60545
	per cap logged	0.1204	0.0246	0.3430	0.0318	0.0872	0.2323	0.2323	0.0872	0.2323	0.2323
	per cap exponentiated	364.57	67.186	1301.1	95.277	272.10	671.29	671.29	101480	671.30	671.30
MSE	total logged	0.0101	0.0010	0.1720	0.0910	0.0114	0.0969	0.0969	0.0114	0.0969	0.0969
	total exponentiated	3.3e+08	6953178	3.4e+09	1.7e+09	1.4e+09	4.3e+10	4.3e+10	1.3e+14	4.3e+10	4.3e+10
	per cap logged	0.0246	0.0008	0.1343	0.0014	0.0129	0.0751	0.0751	0.0129	0.0751	0.0751
	per cap exponentiated	251377	5660.5	2316260	12114	161626	805622	805622	1.2e+10	806296	806296
MAPE	total logged	0.5131	0.2353	3.3665	2.3606	0.8437	2.7340	2.7340	0.8437	2.7340	2.734
	total exponentiated	12.144	2.7184	48.767	32.268	8.5896	24.482	24.482	2668.1	24.481	24.481
	per cap logged	1.4870	0.3127	4.2899	0.4022	1.0991	2.9355	2.9355	1.0991	2.9355	2.9355
	per cap exponentiated	10.937	2.5028	42.107	3.4360	9.0135	22.228	22.228	3474.8	22.226	22.226
RMSPE	total logged	0.8750	0.2839	3.5995	2.6096	1.1251	3.6799	3.6799	1.1251	3.6799	3.6799
	total exponentiated	15.502	3.2731	53.495	36.604	11.172	28.663	28.663	2684.3	28.668	28.668
	per cap logged	1.9166	0.3699	4.5529	0.4820	1.4253	3.4520	3.4520	1.4253	3.4520	3.4520
	per cap exponentiated	13.951	2.9514	45.692	3.9776	12.073	25.690	25.690	3498.0	25.694	25.694

Notes: SP-FE denotes static panel data model using fixed-effect estimator. DP1–3 denote -xtabond and -xtdpd respectively, which represent dynamic panel data model using Arellano-Bond and Arellano-Bover/Blundell-Bond estimators.

exp1 refers to one way to exponentiate logged expenditures using eyf=e^{yf} for the mean of forecast, and eyfstdf=$\sqrt{e^{yfstdf^2}}$ for the standard error of forecast.

exp2 refers to alternative and preferable way to exponentiate logged expenditures using eyf=$e^{yf+(yfstdf^2)/2}$ for the mean of forecast, and eyfstdf=$\sqrt{e^{2yf+yfstdf^2}(e^{yfstdf^2}-1)}$ for the standard error of forecast (Lin, 2012).

Table 3.5: Comparison of forecast performance of time series models versus panel data models using alternative indicators, PHE, logged total and per capita expenditures adjusted by PPP, Canada, within test set 1997–2008

	Series Measure	DE	AS	AD	VAR	SP-FE-exp1	DP1-exp1	DP2-exp1	DP3-exp1	SP-FE-exp2	DP1-exp2	DP2-exp2	DP3-exp2
Theil's U	total logged	0.0204	0.0012	0.0072	0.0043	0.0065	0.0089	0.0043	0.0143	0.0065	0.0089	0.0043	0.0143
	total exponentiated	0.2397	0.0109	0.0465	0.0577	0.0605	0.0833	0.0406	0.1327	0.4783	0.0915	0.0437	0.1460
	per cap logged	0.0301	0.0017	0.0094	0.0064	0.0087	0.0132	0.0051	0.0191	0.0087	0.0132	0.0051	0.0191
	per cap exponentiated	0.2364	0.0111	0.0878	0.0328	0.0653	0.0958	0.0369	0.1319	0.5967	0.1041	0.0401	0.1432
MAE	total logged	0.3873	0.0226	0.1507	0.0739	0.1081	0.1411	0.0687	0.2317	0.1081	0.1411	0.0687	0.2317
	total exponentiated	22766	1427.5	12397	5833.8	9035.9	13508	5515.5	22306	136008	13563	5511.6	22369
	per cap logged	0.3917	0.0221	0.1377	0.0747	0.1071	0.1597	0.0584	0.2313	0.1071	0.1597	0.0584	0.2313
	per cap exponentiated	713.42	43.931	348.86	168.91	224.78	305.76	112.72	427.32	4427.4	305.70	112.63	427.25
MSE	total logged	0.1971	0.0007	0.0275	0.0090	0.0176	0.0326	0.0077	0.0849	0.0176	0.0326	0.0077	0.0849
	total exponentiated	7.4e+08	2375067	2.2e+08	59249140	6.2e+08	1.3e+09	2.5e+08	3.7e+09	8.8e+10	1.3e+09	2.5e+08	3.7e+09
	per cap logged	0.2013	0.0006	0.0230	0.0095	0.0177	0.0390	0.0057	0.0814	0.0177	0.0390	0.0057	0.0814
	per cap exponentiated	709565	2311.7	170903	53974	93128	156303	23505	293149	26444234	156227	23414	293096
MAPE	total logged	3.4534	0.2055	1.3464	0.6588	1.1292	1.5800	0.7474	2.6033	1.1292	1.5800	0.7474	2.6033
	total exponentiated	30.473	2.2880	16.606	7.7000	10.583	13.373	6.7411	21.228	201.06	13.373	6.7444	21.231
	per cap logged	5.0571	0.2923	1.7829	0.9620	1.4271	2.1299	0.7820	3.0847	1.4271	2.1299	0.7820	3.0847
	per cap exponentiated	30.748	2.2332	15.056	7.1125	10.575	15.189	5.7323	21.339	287.71	15.189	5.7326	21.339
RMSPE	total logged	3.9342	0.2339	1.4740	0.8386	1.4633	2.2880	1.0361	3.7014	1.4633	2.2880	1.0361	3.7014
	total exponentiated	33.9577	2.6024	18.473	9.2605	13.128	16.369	8.4508	25.097	203.57	16.368	8.4488	25.101
	per cap logged	5.7479	0.3371	1.9475	1.2419	1.7965	2.6156	1.0134	3.7741	1.7965	2.6156	1.0134	3.7741
	per cap exponentiated	34.229	2.5665	16.713	9.0803	13.376	17.985	7.3081	24.781	290.73	17.986	7.3032	24.784

Notes: SP-FE denotes static panel data model using fixed-effect estimator. DP1,2,3 denote -xtabond, -xtdpdsys, and xtdpd respectively, which further represent dynamic panel data model using Arellano-Bond and Arellano-Bover/Blundell-Bond estimators.

exp1 refers to one way to exponentiate logged expenditures using eyf=e^{yf} for the mean of forecast, and eyfstdf=$\sqrt{e^{yfstdf^2}}$ for the standard error of forecast.

exp2 refers to alternative and preferable way to exponentiate logged expenditures using eyf=$e^{yf+(yfstdf^2)/2}$ for the mean of forecast, and eyfstdf=$\sqrt{e^{2yf+yfstdf^2}(e^{yfstdf^2}-1)}$ for the standard error of forecast (Lin, 2012).

Table 3.6: Combined out-of-sample point forecasts and forecast intervals of per capita THE and PHE in International Dollars for OECD countries in 2015, 2020, and 2025

		THE			PHE				THE			PHE		
		2015	2020	2025	2015	2020	2025		2015	2020	2025	2015	2020	2025
	Upper	5380	6966	9090	3826	5031	6705		7751	10769	14770	4027	5452	7517
Australia	Point	4884	6292	8148	3427	4604	6231	Netherlands	9025	15816	28488	4596	6789	10131
	Lower	4596	6032	7968	2723	3611	4849		6102	8295	11307	3291	4509	6277
	Upper	7326	10099	13937	5548	7573	10404		4750	6110	7889	4228	5608	7864
Austria	Point	6076	8150	10969	4775	6453	8759	New Zealand	4290	5870	8069	3686	5381	7969
	Lower	5366	7432	10366	4169	5689	7841		3346	4344	5667	3057	4506	6318
	Upper	6278	8337	11064	5331	8745	14766		9780	14361	21017	7327	10446	15039
Canada	Point	6101	8347	11444	4911	6073	8426	Norway	7963	11120	15633	6662	9199	12835
	Lower	5155	6895	9279	3510	4383	5768		7390	10918	16336	6476	9136	13078
	Upper	7604	9281	12792	5824	7846	10534		5213	7750	11742	2773	3558	4369
Denmark	Point	6456	9015	12660	5376	7308	9910	Portugal	4224	6186	9168	2226	2669	3148
	Lower	5506	8339	11493	4622	5897	7451		3508	5265	8038	1836	1950	2017
	Upper	5001	6710	9034	3734	5194	7331		5339	8224	13027	4141	7361	11791
Finland	Point	4371	5757	7615	3239	4352	5904	Spain	4653	6855	10311	3495	5230	7861
	Lower	3851	5229	7162	2839	4056	5858		4322	6846	11155	2997	4408	7041
	Upper	6473	8666	11891	5033	7033	9988		5079	6341	8066	4368	5506	6838
Germany	Point	5878	7825	10548	4513	6031	8154	Sweden	4697	5848	7343	3889	4756	5767
	Lower	5387	7303	10134	4173	5610	7740		4210	5351	6900	3314	3766	4183
	Upper	5588	7404	9910	4452	5802	7762		2788	9704	4958	–	–	–
Iceland	Point	4458	5557	7025	3625	4442	5559	Turkey	1863	3149	5375	–	–	–
	Lower	4112	5598	7716	3238	4213	5641		1255	1462	3421	–	–	–
	Upper	5711	7213	8786	4918	6712	9370		5173	6781	9054	4468	6474	8706
Ireland	Point	5067	6776	9055	3820	4999	6668	UK	4857	6582	9036	4256	6072	8468
	Lower	2905	3524	4271	3228	4470	6393		3976	5215	7063	3534	4877	6536
	Upper	5295	7600	11147	4397	6747	10694		11566	15091	19770	5985	8609	12404
Japan	Point	4372	5984	8324	3683	5242	7678	US	10413	13287	17049	5536	7812	11046
	Lower	4127	6015	8943	3781	5830	9268		9579	12598	16773	5050	7330	10675
	Upper	3899	6208	9861	2270	3693	5710							
South Korea	Point	3276	5184	8206	2018	3429	5709							
	Lower	2730	4216	6492	1480	2348	3500							

Notes: Canada's PHE/THE point forecast in 2015 is 80%, which is higher than historical average as 70% during 2000–2012. Recall this is the forecast combined from four smoothing and time series models. The individual forecast of PHE/THE provided by individual model DE, HW, AD, and VAR respectively is 74%, 97%, 71%, and 80%. Obviously, the final combined 80% comes from the VAR and especially the HW model, which corresponds to HW's high sensitivity to the end points of data used to forecast.

Similarly, the Netherlands' forecasts are much larger than the US' forecasts, because that the individual forecasts from the statistical smoothing models especially HW model for the Netherlands are much larger than those for the US. Also, the Netherlands has a higher rate of growth in PHE because of their reforms in 2005–2006, which gets projected forward.

Table 3.7: Ranks of combined point forecasts from time series data models, per capita, International Dollars, OECD countries, 2015, 2020, and 2025

	THE						PHE						PHE/THE					
Rank	2015		2020		2025		2015		2020		2025		2015		2020		2025	
1	USA	10413	NLD	15816	NLD	28488	NOR	6662	NOR	9199	NOR	12835	GBR	0.88	GBR	0.92	NZL	0.99
2	NLD	9025	USA	13287	USA	17049	USA	5536	USA	7812	USA	11046	NZL	0.86	NZL	0.92	GBR	0.94
3	NOR	7963	NOR	11120	NOR	15633	DNK	5376	DNK	7308	NLD	10131	JPN	0.84	JPN	0.88	JPN	0.92
4	CHE	7079	CHE	9327	DNK	126604	CAN	4911	NLD	6789	DNK	9910	NOR	0.84	NOR	0.83	NOR	0.82
5	DNK	6456	DNK	9015	CHE	12281	AUT	4775	AUT	6453	AUT	8759	DNK	0.83	SWE	0.81	AUT	0.80
6	CAN	6101	CAN	8347	CAN	11444	NLD	4596	CAN	6073	GBR	8468	SWE	0.83	DNK	0.81	ISL	0.79
7	AUT	6076	AUT	8150	AUT	10969	DEU	4513	GBR	6072	CAN	8426	ISL	0.81	ISL	0.80	SWE	0.79
8	DEU	5878	DEU	7825	DEU	10548	GBR	4256	DEU	6031	DEU	8154	CAN	0.80	AUT	0.79	DNK	0.78
9	BEL	5490	BEL	7441	ESP	10311	SWE	3889	NZL	5381	NZL	7969	AUT	0.79	DEU	0.77	FIN	0.78
10	IRE	5067	ESP	6855	BEL	10186	IRE	3820	JPN	5242	ESP	7861	DEU	0.77	ESP	0.76	DEU	0.77
11	AUS	4884	IRE	6776	PRT	9168	NZL	3686	ESP	5230	JPN	7678	IRE	0.75	FIN	0.76	AUS	0.76
12	GBR	4857	GBR	6582	IRE	9055	JPN	3683	IRE	4999	IRE	6668	ESP	0.75	IRE	0.74	ESP	0.76
13	SWE	4697	AUS	6292	GBR	9036	ISL	3625	SWE	4756	AUS	6232	FIN	0.74	AUS	0.73	IRE	0.74
14	ESP	4653	PRT	6186	JPN	8324	ESP	3495	AUS	4604	FIN	5904	AUS	0.70	CAN	0.73	CAN	0.74
15	ISL	4458	JPN	5984	KOR	8206	AUS	3427	ISL	4442	SWE	5767	KOR	0.62	KOR	0.66	KOR	0.70
16	JPN	4372	NZL	5870	AUS	8148	FIN	3239	FIN	4352	KOR	5709	USA	0.53	USA	0.59	USA	0.65
17	FIN	4371	SWE	5848	NZL	8069	PRT	2226	KOR	3429	ISL	5559	PRT	0.53	PRT	0.43	NLD	0.36
18	NZL	4290	FIN	5757	FIN	7615	KOR	2018	PRT	2669	PRT	3148	NLD	0.51	NLD	0.43	PRT	0.34
19	PRT	4224	ISL	5557	SWE	7343	–	–	–	–	–	–	–	–	–	–	–	–
20	KOR	3276	KOR	5184	ISL	7025	–	–	–	–	–	–	–	–	–	–	–	–
21	TUR	1863	TUR	3149	TUR	5375	–	–	–	–	–	–	–	–	–	–	–	–

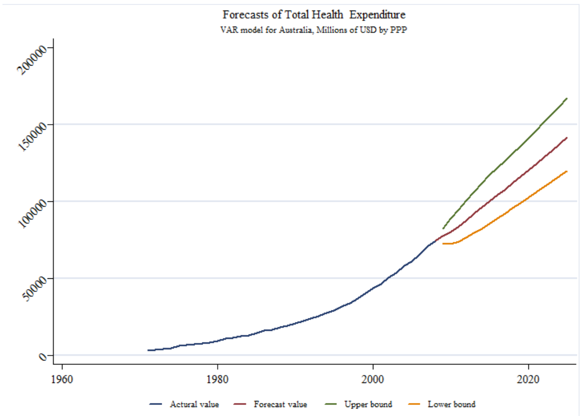
Notes: According to OECD country abbreviations, NLD—the Netherlands; NOR—Norway; CHE—Switzerland; DNK—Denmark; CAN—Canada; AUT—Austria; DEU—Germany; BEL—Belgium; IRE—Ireland; AUS—Australia; GBR—Great Britain; SWE—Sweden; ESP—Spain; ISL—Iceland; JPN—Japan; FIN—Finland; NZL—New Zealand; PRT—Portugal; KOR—South Korea; TUR—Turkey.

Table 3.8: Large and systematic discrepancies among international databases

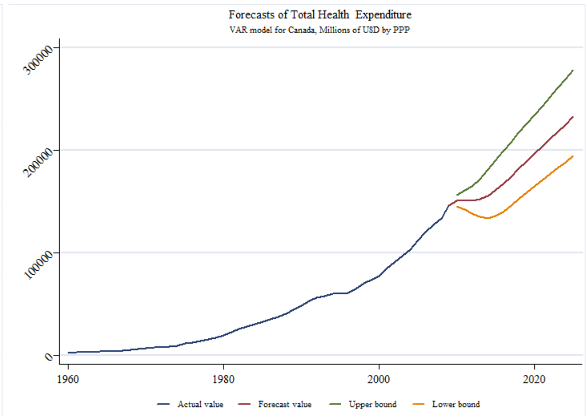
PHE (Billions NCU)	Country	Switzerland	Denmark													
	Year	2008	2002	2003	2004	2005	2006	2007	2008	2009						
	OECD	35	108	113	120	128	137	143	151	162						
	WHO	34	100	104	110	116	125	132	138	149						
	IMF	11	96	99	105	111	119	127	135	–						
OOP (Billions NCU)		Japan														
	Year	1995	1996	1997	1998	1998	1999	2001	2002	2003	2004	2005	2006	2007	2008	
	OECD	4,775	5,020	5,468	5,775	5,774	5,959	5,977	6,099	6,540	6,543	6,350	7,059	6,807	6,774	
	WHO	5,198	5,544	6,059	6,446	6,432	6,524	6,517	6,763	6,079	6,112	5,843	6,207	6,115	6,098	
GDP (Billions NCU)		South Korea														
	Year	1995	1996	1997	1998	1999	2000	...	2009							
	OECD	409,654	460,953	506,314	501,027	549,005	603,236	...	1,065,037							
	WHO	415,773	467,645	511,990	504,659	551,983	603,236	...	1,063,059							
	WB	398,838	448,596	491,135	484,103	529,500	603,236	...	1,063,059							
	UN	409,654	460,953	506,314	501,027	549,005	603,236	...	–							
PHE % GGE		Switzerland	Luxembourg													
	Year	2007	2002	2003												
	WHO	19.54	14.77	14.02												
	WB	19.54	14.77	14.02												
	IMF	6.00	11.28	11.45												
PHE % THE		Belgium														
	Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
	OECD	76.80	78.18	75.38	74.80	74.58	74.61	75.41	73.80	74.78	76.02	75.93	73.87	73.51	74.96	75.10
	WHO	68.67	70.68	68.21	67.82	67.72	67.53	68.70	67.18	70.51	71.42	72.03	72.77	67.95	66.78	68.35
	WB	68.67	70.68	68.21	67.82	67.72	67.53	68.70	67.18	70.51	71.42	72.03	72.77	67.95	66.78	68.35
		Chile														
	Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	...	2009				
	OECD	48.18	47.17	47.14	48.11	49.86	52.10	53.54	54.51	38.81	...	47.38				
	WHO	38.27	36.63	37.34	38.28	39.90	41.64	42.95	43.76	38.81	...	47.37				
	WB	48.18	47.17	47.14	48.11	49.86	52.10	53.54	54.51	38.81	...	46.79				
		Denmark														
	Year	1995	...	2003	2004	2005	2006	2007	2008	2009						
	OECD	82.52	...	84.55	84.27	84.48	84.64	84.40	84.66	85.04						
	WHO	82.52	...	79.75	79.16	79.35	79.99	80.21	80.15	80.09						
	WB	82.52	...	79.75	79.16	79.35	79.99	80.21	80.15	80.09						
		Finland														
	Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
	OECD	71.68	71.62	72.11	72.47	71.46	71.26	71.97	72.46	74.56	74.97	75.39	74.85	74.36	74.45	74.71
	WHO	82.52	82.42	82.28	81.99	82.21	82.43	82.67	82.94	79.75	79.16	79.35	79.99	80.21	80.15	80.09
	WB	72.01	71.91	72.18	71.81	71.48	71.06	71.83	72.31	68.27	69.01	69.47	70.16	70.15	70.70	72.05
		Australia														
	Year	1995	...	2003	2004	2005	2006	2007	2008							
	OECD	65.78	...	66.11	66.68	66.89	66.59	67.51	67.99							
	WHO	65.78	...	66.11	66.68	66.89	66.58	67.51	70.10							
	WB	65.78	...	64.54	64.56	64.47	64.16	65.40	65.40							

Notes: The currency unit for PHE, OOP, GDP is the current price.

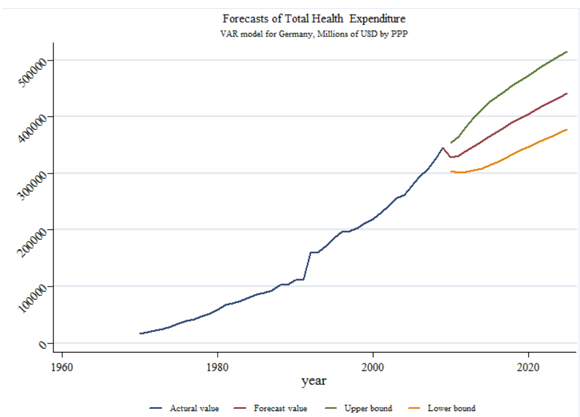
Figure 3.1: Forecast intervals of THE in total millions international dollar from time series VAR models for OECD countries—Australia, Canada, Germany, and Iceland, 2010–2025, in different scales on the vertical axes



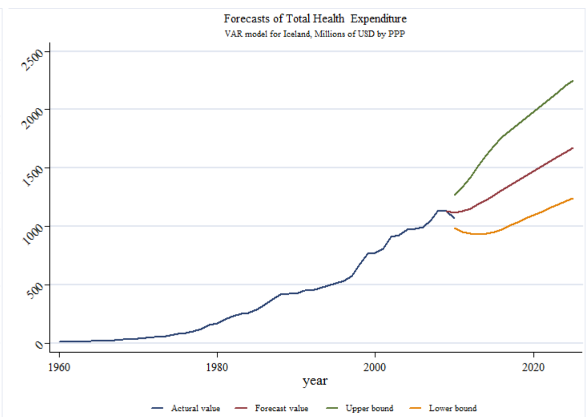
(a) Australia



(b) Canada

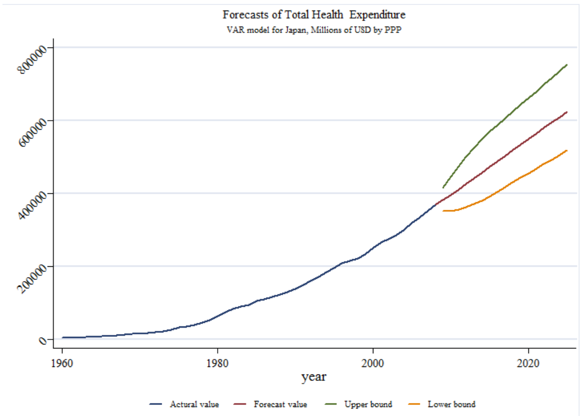


(c) Germany

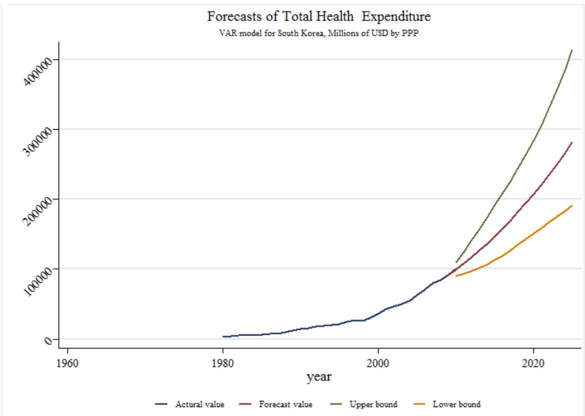


(d) Iceland

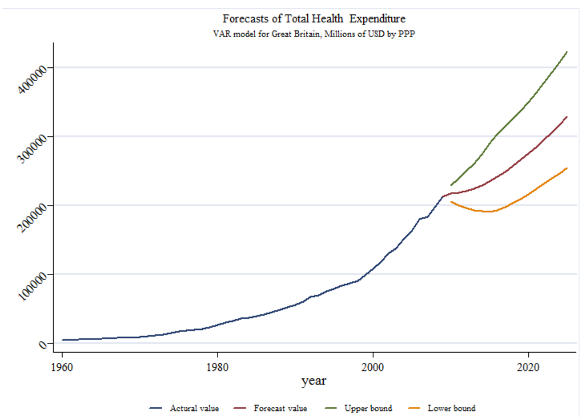
Figure 3.2: Forecast intervals of THE in total millions international dollar from time series VAR models for OECD countries—Japan, South Korea, Great Britain, and United States, 2010–2025, in different scales on the vertical axes



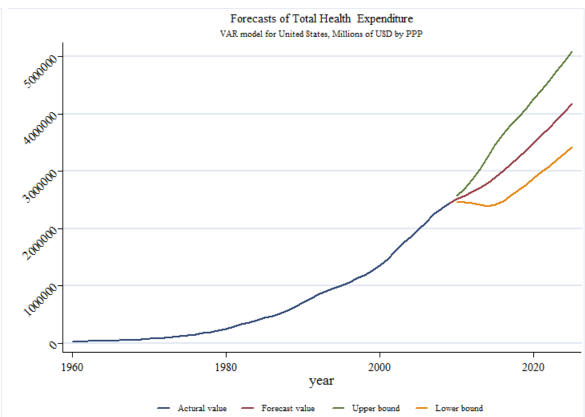
(a) Japan



(b) South Korea

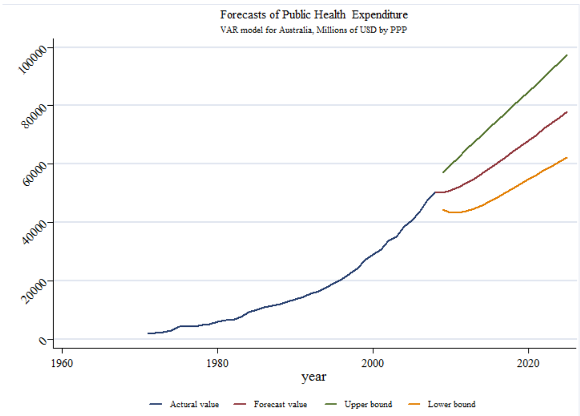


(c) Great Britain

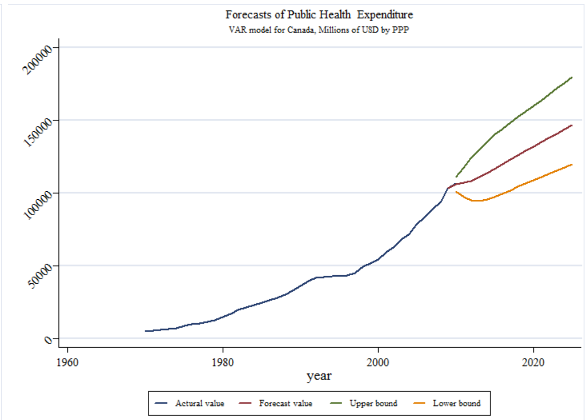


(d) United States

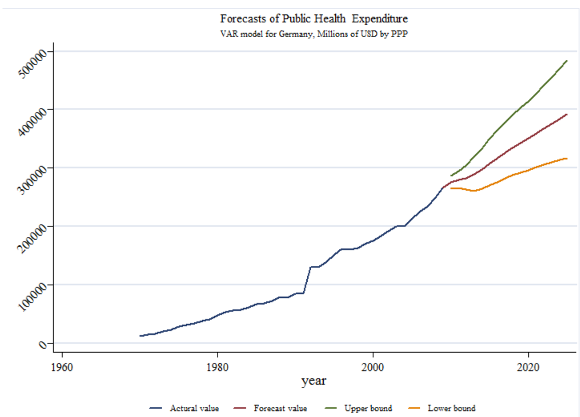
Figure 3.3: Forecast intervals of PHE in total millions international dollar from time series VAR models for OECD countries—Australia, Canada, Germany, and Iceland, 2010–2025, in different scales on the vertical axes



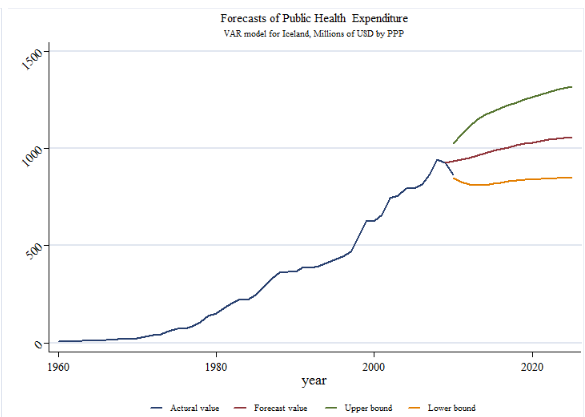
(a) Australia



(b) Canada

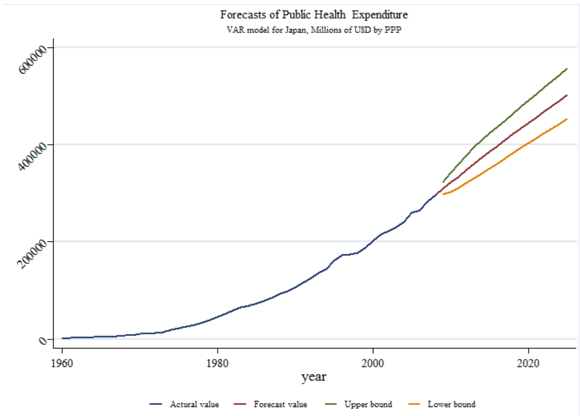


(c) Germany

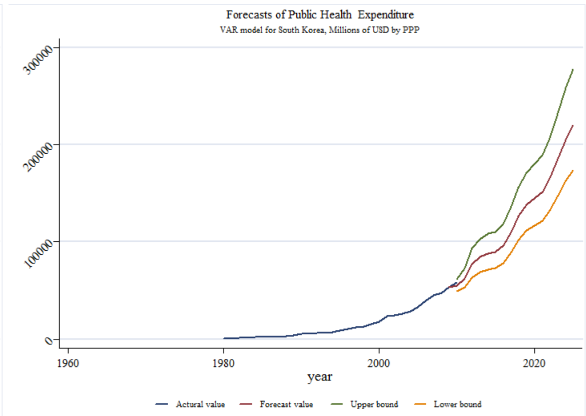


(d) Iceland

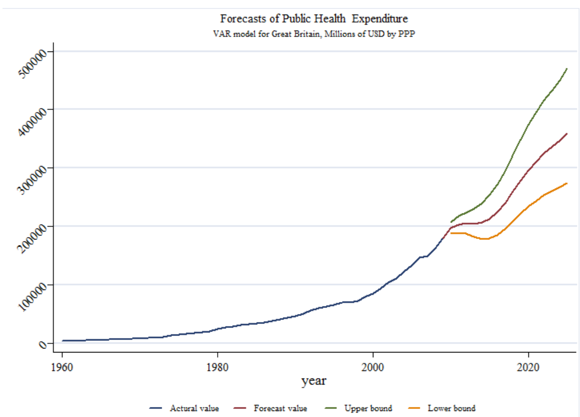
Figure 3.4: Forecast intervals of PHE in total millions international dollar from time series VAR models for OECD countries—Japan, South Korea, Great Britain, and United States, 2010–2025, in different scales on the vertical axes



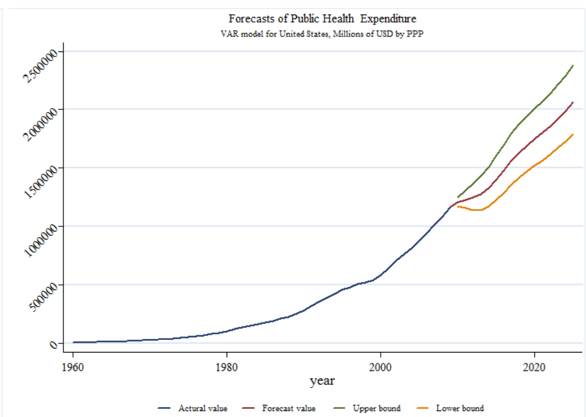
(a) Japan



(b) South Korea



(c) Great Britain



(d) United States

Conclusion

This dissertation consists of three studies that use different ways of understanding the world. The pure theoretical study constructs a rational and anonymous social ordering function satisfying a relaxed IIA to prove a possibility theorem. The applied theoretical study models a two-sector macroeconomic model of the economy, and simulates the equilibrium shift and transitional dynamics caused by a policy to invest in health (by reallocating labor from non-health to the health sector). The empirical study examines the performance of various forecasting models (e.g., times series versus panel), and based on the better-performing ones provides forecasts of health expenditures for OECD countries in the short and medium run.

In the social choice chapter, we work in a higher-dimensional economic environment where an indifference hypersurface is the level set of a utility function representing its preference relation. We introduce the concept of curvature that describes the shape of an indifference hypersurface. The IIA assumption is weakened by adding information about the indifference hypersurfaces via curvature. We show that using such information it is possible to construct a rational, anonymous social ordering function that satisfies this weaker IIA condition. Our curvature-based social ordering function coincides with the Rawlsian difference principle under certain circumstances. The importance of this pure theorem is to show that the conditions for democracy can be weaker than many think.

In the health capital chapter, we have three major findings. First, a policy that invests in health (here modeled by a reallocation of labor from the manufacturing to the health sector) improves health status and welfare in the long run, but harms the economic growth in both short and long term. The relative sizes of these competing effects depend on the specific health features of the country. Such long-run welfare and growth effects we find are consistent with existing theoretical literature. Our findings challenge the policy rationales of World Bank (1993) and World Health Organization (2001) in the sense that good health, though improves welfare, increases neither substantially the economic productivity of workers nor the economic growth rate of countries. Second, under the full range of plausible values for the relative weight households place on health versus consumption, households are made better off by the health investment policy as long as the effectiveness of health investment in producing health is sufficiently large. Households can be worse off by such a policy only if both health investment is not productive in generating health and households value health relatively little relative to consumption. Third, unlike much of the existing literature, this chapter analyzes an investment in the public health sector, which corresponds better to the institutional arrangements of developed countries who have a publicly funded health sector like Canada. For these countries where the productivity-enhancing effect of health is small, investments in health look very promising given their welfare gains for households in the long run. However, such investments slow down the economic growth in both the short and long run. Fourth, our model shows the robustness of the Van Zon and Muysken (2005) and Hall and Jones (2007)' findings about the welfare and growth effects of a health investment policy.

In the health expenditure forecasting chapter, our results indicate that, contrary to Getzen and Poullier (1992), complicated econometric (esp. static) panel data models perform worse than simpler statistical (e.g., smoothing) and econometric time-series data models for forecasting. This finding is, however, consistent with empirical findings from the broader literature on economic forecasting (Armstrong, 1978; McNees, 1986). Therefore, to forecast, more complicated econometric models — except VAR — generally do not do better than simpler ones such as exponential smoothing. This also implies that a (usually complex) model that better fits historical data does not guarantee accurate and precise post-sample forecasts. This chapter also finds that some methods perform more accurately for short horizons (e.g., exponential smoothing) while others are more appropriate for medium horizons (e.g., ARIMA, multivariate regressions such as VAR). This is also consistent with studies from the broader literature on forecasts (Fildes and Makridakis, 1995; Fildes *et al.*, 1998). This chapter also contributes to the call articulated in Astolfi *et al.* (2012) for more rigorous methods, for more transparency, and for better assessment of performance. Forecasting health expenditures are crucial for policy applications required by governmental organizations and central banks. Readers should carefully distinguish the objective of simply getting accurate forecasts of future spending from modeling and understanding determinants of expenditure growth. Modeling and understanding the determinants of expenditure growth require considerably more data and effort, and may still do worse with pure forecasting. This chapter confirms this.

Bibliography

Aghion, P., Caroli, E., and García-Peñalosa, C. (1999). Inequality and Economic Growth: The Perspective of the New Growth Theories. *Journal of Economic Literature*, **37**, 1615–1660.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58**, 277–298.

Arellano, M. and Bover, O. (1995). Another Look at Instrumental Variables Estimation of Error Components Models. *Journal of Econometrics*, **68**, 29–51.

Armstrong, J. (1978). Forecasting With Econometric Methods: Folklore Versus Facts. *Journal of Business*, **51**, 549–564.

Armstrong, J. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic.

Astolfi, R., Lorenzoni, L., and Oderkirk, J. (2012). Informing Policy Makers About Future Health Spending: A Comparative Analysis of Forecasting Methods in OECD Countries. *Health Policy*, **107**, 1–10.

Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Wiley, New York.

Bank of England (2015). Inflation Report November 2005. http://www.bankofengland.co.uk/publications/Documents/inflationreport/ir05nov.pdf.

Barro, R. and Sala-i Martin, X. (1995). *Economic Growth*. McGraw-Hill, Inc.

Barros, P. (1998). The Black-Box of Health Care Expenditure Growth Determinants. *Health Economics*, **7**, 533–544.

Bloom, D., Canning, D., and Sevilla, J. (2004). The Effect of Health on Economic Growth: A Production Function Approach. *World Development*, **32**, 1–13.

Blundell, R. and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, **87**, 115–143.

Box, G., Jenkins, G. M., and Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken.

Brauer, F., van den Driessche, P., Wu, J., Allen, L., Bauch, C., and Castillo-Chavez, C. (2008). *Mathematical Epidemiology*. Springer.

Cameron, A. and Trivedi, P. (2010). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Campbell, D. (1992). *Equity, Efficiency, and Social Choice*. Clarendon Press, Oxford.

Campbell, D. and Kelly, J. (2007). Social Welfare Functions That Satisfy Pareto, Anonymity, and Neutrality, but not Independence of Irrelevant Alternatives. *Social Choice and Welfare*, **29**, 69–82.

Campbell, D. and Kelly, J. (2009). Uniformly Bounded Information and Social Choice. *Journal of Mathematical Economics*, **45**, 415–421.

Carnot, N., Koen, V., and Tissot, B. (2011). *Economic Forecasting and Policy*. Palgrave Macmillan, New York.

Centers for Medicare & Medicaid Services (1991). The Long-Term Projection Assumptions for Medicare and Aggregate National Health Expenditures.

Centers for Medicare & Medicaid Services (2000). Technical Panel Reports. Centers for Medicare & Medicaid Services.

Centers for Medicare & Medicaid Services (2004). Technical Panel reports. Centers for Medicare & Medicaid Services.

Centers for Medicare & Medicaid Services (2011). Projected Medicare Expenditures Under an Illustrative Scenario With Alternative Payment Updates to Medicare Providers.

Clemen, R. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, **5**, 559–584.

Clements, M. and Hendry, D. (2004). *A Companion to Economic Forecasting*. Basil Blackwell, Oxford.

Clements, M. and Hendry, D. (2011). *The Oxford Handbook of Economic Forecasting*. Oxford University Press, New York.

Culyer, A. J. (1988). Health Care Expenditures in Canada: Myth and Reality; Past and Future. Canadian Tax Paper No. 82. *Canadian Tax Foundation, Toronto.*

Debreu, G. (1972). Smooth Preferences. *Econometrica*, **40**, 603–615.

Denton, F., Gafni, A., and Spencer, B. (2002). Exploring the Effects of Population Change on the Costs of Physician Services. *Journal of Health Economics*, **21**, 781–803.

Di Matteo, L. (2010). The Sustainability of Public Health Expenditures: Evidence From the Canadian Federation. *The European Journal of Health Economics*, **11**, 569–584.

Diebold, F. X. and Kilian, L. (2001). Measuring Predictability: Theory and Macroeconomic Applications. *Journal of Applied Econometrics*, **16**, 657–669.

Dierker, H. (1975). Smooth Preferences and the Regularity of Equilibria. *Journal of Mathematical Economics*, **2**, 43–62.

Dolado, J., Jenkinson, T., and Sosvilla-Rivera, S. (1990). Cointegration and Unit Roots: A Survey. *Journal of Economic Surveys*, **4**, 249–273.

Elliott, G. and Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, **46**, 3–56.

Elliott, G., Rothenberg, T., and Stock, J. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, **64**, 813–836.

European Union (2008). The 2009 Ageing Report: Underlying Assumptions and Projection Methodologies. `http://ec.europa.eu/economy_finance/publications/publication13782_en.pdf`.

European Union (2009). Long-Term Sustainability of Public Finances for a Recovering Economy. `http://ec.europa.eu/economy_finance/publications/publication15996_en.pdf`.

Evans, R., McGrail, K., Morgan, M., and Hertzman, C. (2001). APOCALYPSE NO: Population Aging and the Future of Health Care System. *Canadian Journal on Aging*, **20**, 160–191.

Fanti, L., Gori, L., and Tramontana, F. (2014). Endogenous Lifetime, Accidental Bequests and Economic Growth. *Decisions in Economics and Finance*, **37**, 81–98.

Fildes, R. and Makridakis, S. (1995). The Impact of Empirical Accuracy Studies on Time-Series Analysis and Forecasting. *International Statistical Review*, **63**, 289–308.

Fildes, R., Hibon, M., Makridakis, S., and Meade, N. (1998). Generalising About Univariate Forecasting Methods: Further Empirical Evidence. *International Journal Of Forecasting*, **14**, 339–358.

Fleurbaey, M., Suzumura, K., and Tadenuma, K. (2005). Arrovian Aggregation in Economic Environments: How Much Should We Know About Indifference Surfaces? . *Journal of Economic Theory*, **124**, 22–44.

Gerdtham, U.-G. and Jonsson, B. (2000). International Comparison of Health Expenditure. In A. Culyer and J. P. Newhouse, editors, *Handbook of Health Economics*. Elsevier.

Gerdtham, U.-G., Sogaard, J., Jonsson, B., and Andersson, F. (1992). A Pooled Crossed Section Analysis of the Health Expenditure of the OECD Countries. In

P. Zweifel and H. Frech, editors, *Health Economics Worldwide*. Kluwer Academic Publishers, Dordrecht.

Getzen, T. (2000). Forecasting Health Expenditures: Short, Medium, and Long Term. *Journal of Health Care Finance*, **26**, 56–72.

Getzen, T. (2006). Aggregation and the Measurement of Health Care Costs. *Health Service Research*, **41**, 1938–1954.

Getzen, T. (2007). Modelling Long Term Healthcare Cost Trends. *The Society of Actuaries*.

Getzen, T. and Poullier, J.-P. (1992). International Health Spending Forecasts: Concepts and Evaluation. *Social Science and Medicine*, **34**, 1057–1068.

Gong, L., Li, H., and Wang, D. (2012). Health Investment, Physical Capital Accumulation, and Economic Growth. *China Economic Review*, **23**, 1104–1119.

Granger, C. (1989). *Forecasting in Business and Economics*. Academic Press, San Diego.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, **80(2)**, 223–255.

Hall, R. and Jones, C. (2007). The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics*, **122(1)**, 39–72.

Halliday, T., He, H., and Zhang, H. (2014). Health Investment Over the Life-Cycle. IZA Discussion Paper No. 4482 . `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1490504`.

Harris, R. and Tzavalis, E. (1999). Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed. *Journal of Econometrics*, **91**, 201–226.

Hayashi, T. (2008). A Note on Small Income Effects. *Journal of Economic Theory*, **139**, 2360–379.

Heckman, J. (2000). Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. *Quarterly Journal of Economics*, **115**, 45–97.

Hendry, D. and Hubrich, K. (2012). Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate. *Journal of Business and Economic Statistics*, **29**, 216–227.

Holden, K., Peel, D., and Thompson, J. (1990). *Economic Forecasting: An Introduction*. Cambridge University Press, Cambridge.

Im, K., Pesaran, M., and Shi, Y. (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, **115**, 53–74.

Inada, K. (1964). On the Economic Welfare Function. *Econometrica*, **32**, 316–338.

Jones, C. (2013). Life and Growth: NBER Reporter 2013 Number 2: Research Summary. *The National Bureau of Economic Research*.

Kalai, E., Muller, E., and Satterthwaite, M. (1979). Social Welfare Functions When Preferences Are Convex, Strictly Monotonic, and Continuous. *Public Choice*, **34**, 87–97.

Knowles, S. and Owen, P. D. (1997). Education and Health in an Effective-labor Empirical Growth Model. *The Economic Record*, **73**, 314–328.

Lago-Peñas, S., Cantarero-Prieto, D., and Blázquez-Fernández, C. (2013). On the Relationship Between GDP and Health Care Expenditure: A New Look. *Economic Modelling*, **32**, 124–129.

Le Breton, M. and Weymark, J. A. (1996). An Introduction to Arrovian Social Welfare Functions on Economic and Political Domains. In N. Schofield, editor, *Collective Decision-Making: Social Choice and Political Economy*. Kluwer Academic Publishers, Boston.

Levin, L., Lin, C.-F., and Chu, S. (2002). Unit Root Tests In Panel Data: Asymptotic and Finite-Sample Properties. *Journal of Econometrics*, **108**, 1–24.

Lin, K.-P. (2012). Time Series Analysis and Forecasting. `http://web.pdx.edu/~crkl/ec572/ec572-1.htm`.

Lorenzoni, L., Belloni, A., and Sassi, F. (2014). Health-Care Expenditure and Health Policy in the USA Versus Other High-Spending OECD Countries. *The Lancet*, **384**, 83–92.

Makridakis, S. and Winkler, R. (1983). Average of Forecasts: Some Empirical Results. *Management Science*, **29**, 987–996.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., and Simmons, L. (1993). The M2-Competition: A Real Time Judgmentally Based Forecasting Study (With Comments). *International Journal of Forecasting*, **9**, 5–30.

Makridakis, S., Wheelwright, S., and Hyndman, R. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons, Hoboken.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*. Oxford University Press.

Mayston, D. (1980). Where Did Prescriptive Welfare Economics Go Wrong? In D. Currie and W. Peters, editors, *Contemporary Economic Analysis, Vol.2*. Croom Helm, London.

McNees, S. (1986). Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts. *Journal of Business and Economic Statistics*, **4**, 5–15.

Narayan, P. and Popp, S. (2012). A Nonlinear Approach to Testing the Unit Root Null Hypothesis: An Application to International Health Expenditures. *Applied Economics*, **44**, 163–175.

NBER (1966). *Long-Rang Economic Projection: A Report of the National Bureau of Economic Research. Conference on Research in Income and Wealth.* Princeton University Press, Princeton.

Newhouse, J. (1977). Medical Care Expenditures: A Cross-National Survey. *Journal of Human Resources*, **12**, 115–152.

Nordhaus, W. (2002). The Health of Nations: The Contribution of Improved Health to Living Standards. *NBER Working Paper No.*, **2**.

OECD (2003). Projecting OECD Health and Long-Term Care Expenditures: What Are the Main Drivers? OECD Economics Department Working Papers, No. 477. *Paris and Washington, D.C.*

OECD (2011). A System of Health Accounts. http://www.oecd-ilibrary.org/social-issues-migration-health/a-system-of-health-accounts_9789264116016-en.

OECD (2015). OECD Health Data. http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT.

OECD Statistics (2014). OECD Real GDP Growth Rate. http://stats.oecd.org/.

Panopoulou, E. and Pantelidis, T. (2011). Convergence in Per Capita Health Expenditures and Health Outcomes in the OECD Countries. *Applied Economics*, **44**, 3909–3920.

Parkin, D., McGuire, A., and Yule, B. (1987). Aggregate Health Expenditures and National Income: Is Health Care a Luxury Good? *Journal of Health Economics*, **6**, 109–127.

Parkin, D., McGuire, A., and Yule, B. (1989). What Do International Comparisons of Health Expenditures Really Show? *Community Medicine*, **11**, 116–123.

Pazner, E. A. and Schmeidler, D. (1978). Egalitarian-Equivalent Allocations: A New Concept of Economic Equity. *The Quarterly Journal of Economics*, **92**, 671–687.

Philips, P. and Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, **75**, 335–346.

Ramsey, F. (1928). A Mathematical Theory of Saving. *Economic Journal*, **38**, 543–559.

Redekop, J. (1991). Social Welfare Functions on Restricted Economic Domains. *Journal of Economic Theory*, **53**, 396–427.

Ried, W. (1996). Willingness to Pay and Cost of Illness for Changes in Health Capital Depreciation. *Health Economics*, **5**, 447–468.

Romer, P. (1990). Endogenous Technical Change. *Journal of Political Economy*, **98**, S71–S102.

Samuelson, P. (1941). The Stability of Equilibrium: Comparative Statics and Dynamics. *Econometrica*, **9**, 97–120.

Samuelson, P. A. (1980). *Economics: An Introductory Analysis*. McGraw-Hill, New York.

Scarth, W. (2014). *Macroeconomics: The Development of Modern Methods for Policy Analysis*. Edward Elgar Publishing, Cheltenham.

Sen, A. (1986). Social Choice Theory. In K. Arrow and M. Intriligator, editors, *Handbook of Mathematical Economics, Vol.3*. North-Holland.

Sims, C. (1972). Money, Income, and Causality. *The American Economic Review*, **62**, 540–552.

Smale, S., Hirsch, M., and Devaney, R. (2012). *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Academic Press.

Stanford Encyclopedia of Philosophy (2013). Social Choice Theory. `http://plato.stanford.edu/entries/social-choice/#ArrHisInf`.

Statistics Canada (2014). Jobs Report: The State of the Canadian Labour Market. `http://www.budget.gc.ca/2014/docs/jobs-emplois/pdf/jobs-emplois-eng.pdf`.

Sydsæter, K., Hammond, P., Seierstad, A., and Strøm, A. (2005). *Further Mathematics for Economic Analysis*. Prentice Hall.

Theil, H. (1966). *Applied Economic Forecasting*. North-Holland, Amsterdam.

Thorpe, J. (1979). *Elementary Topics in Differential Geometry*. Springer, New York.

Tompa, E. (2002). The Impact of Health on Productivity: Empirical Evidence and Policy Implication. In A. Sharpe, F. St-Hilaire, and K. Banting, editors, *The Review of Economic Performance and Social Progress: Toward a Social Understanding of Productivity*. The Institute for Research on Public Policy.

Turnovsky, S., Atolia, M., and Chatterjeeb, S. (2010). How Misleading is Linearization? Evaluating the Dynamics of the Neoclassical Growth Model. *Journal of Economic Dynamics and Control*, **34**, 1550–1571.

Van Zon, A. and Muysken, J. (2001). Health and Endogenous Growth. *Journal of Health Economics*, **20**, 169–185.

Van Zon, A. and Muysken, J. (2005). Health as a Principle Determinant of Economic Growth. In G. Lopez-Casasnovas, B. Rivera, and L. Currais, editors, *Health and Economic Growth: Findings and Policy Implications*. The MIT Press.

Westerlund, J. (2007). Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics*, **69**, 709–748.

World Bank (1993). *World Development Report 1993: Investing in Health*. Oxford University Press, Washington, D.C.

World Bank (2015). Gross Domestic Investment as Percentage of GDP . `http://data.worldbank.org/indicator/NE.GDI.TOTL.ZS`.

World Health Organization (2001). *Macroeconomics and Health: Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health*. World Health Organization, Geneva.