FISCAL POLICY FOR HEALTH POLICY MAKERS

Robert Gillingham

March 2014
FISCAL POLICY FOR HEALTH POLICY MAKERS

Robert Gillingham

March 2014
Health, Nutrition, and Population (HNP) Discussion Paper

This series is produced by the Health, Nutrition, and Population (HNP) family of the World Bank's Human Development Network (HDN). The papers in this series aim to provide a vehicle for publishing preliminary results on HNP topics to encourage discussion and debate. The findings, interpretations, and conclusions expressed in this paper are entirely those of the author(s) and should not be attributed in any manner to the World Bank, to its affiliated organizations, or to members of its Board of Executive Directors or the countries they represent. Citation and the use of material presented in this series should take into account this provisional character.

For information regarding the HNP Discussion Paper Series, please contact Martin Lutalo at mlutalo@worldbank.org or Erika Yanick at eyanick@worldbank.org.

© 2014 The International Bank for Reconstruction and Development / The World Bank
1818 H Street, NW, Washington, DC 20433
All rights reserved.
Abstract: This paper summarizes the basic principles that should form the basis of fiscal policy. These principles encompass decisions on the functions of government, its spending, and the financing of its spending that affect economic growth, employment, inflation, and economic welfare. Although the principles are broadly applicable, it is especially important that health policy makers understand them. Ensuring access to health care is one of society’s — and therefore the government’s — most important goals. In meeting this goal, policy makers must be cognizant of fiscal realities — what they can reasonably expect government to achieve in the health sector and at what cost. Resources are limited, and many valuable programs in both the private and public sectors vie for them. Allocating these resources efficiently is of paramount importance, with implications for what the government does and how it finances its activities. The absolute level and share of government resources allocated to the health sector will depend on a variety of factors, but the bottom line is that health programs must compete with other government programs for scarce resources to ensure that these resources are put to their best use.

Keywords: Health, fiscal policy for health, public finance, revenue mobilization, expenditure decision making.

Disclaimer: The findings, interpretations, and conclusions expressed in the paper are entirely those of the author, and do not represent the views of the World Bank, its Executive Directors, or the countries they represent.

Correspondence Details: Rafael Cortez, The World Bank Group, 1818 H Street, NW, Washington, DC 20433; Telephone: 202-458-8707; fax: 202-522-3234; e-mail: rcortez@worldbank.org; website: www.worldbank.org/hnp.
# Table of Contents

I. INTRODUCTION AND KEY ISSUES.................................................................................. 1

II. SETTING EXPENDITURE PRIORITIES.......................................................................... 2

  2.1. THE PROPER SCOPE OF GOVERNMENT............................................................... 5

  2.2. FOCUS ON GOVERNMENT FUNCTIONS AND PROGRAMS.................................. 7

  2.3. MAKE INFORMED JUDGMENTS........................................................................... 9

  2.4. AVOID PITFALLS.................................................................................................. 10

III. LONG-TERM EXPENDITURE CHALLENGES............................................................. 11

  3.1. PUBLIC PENSIONS............................................................................................. 12

  3.2. HEALTH CARE.................................................................................................. 13

IV. RAISING REVENUE.................................................................................................... 14

  4.1. CONSUMPTION TAXES....................................................................................... 16

  4.2. INCOME TAXES................................................................................................ 20

  4.3. WEALTH TAXES................................................................................................ 23

  4.4. ADDITIONAL REVENUE POLICY ISSUES......................................................... 24

  4.5. NONTAX REVENUE............................................................................................ 25

  4.6. VARIATIONS IN REVENUE SOURCES BY DEVELOPMENT LEVEL...................... 26

V. FISCAL SUSTAINABILITY............................................................................................... 27

VI. IMPLICATIONS FOR HEALTH POLICY.................................................................... 31

  1.1. ALLOCATIVE EFFICIENCY................................................................................... 38

  1.2. PRODUCTIVE EFFICIENCY.................................................................................. 39

  1.3. TYPES OF GOVERNMENT INTERVENTION....................................................... 39

ANNEX 2. THE ROLE OF THE STATE IN THE ECONOMY.............................................. 41

  2.1. PARETO EFFICIENCY........................................................................................ 41

  2.2. REDISTRIBUTION............................................................................................... 44

ANNEX 3. EXCESS BURDEN OF TAXATION................................................................. 46
ACKNOWLEDGMENTS

This report was prepared by Robert Gillingham (Consultant, HDNHE) as part of the Human Development Network Health Economy Program, led by Rafael Cortez (Task Team Leader, Senior Economist, HDNHE). The program work was conducted under the overall supervision of Nicole Klingen (Sector Manager, HDNHE) and Timothy Evans (Sector Director, HDNHE) and funded by the Rockefeller Foundation.

The author is grateful for comments and inputs on several drafts of the paper from Jack Langenbrunner, George Schieber and Rafael Cortez. The author also benefited greatly from discussions with and comments from Cheryl Cashin, Lisa Fleisher, Michael Keen, Joseph Kutzin, Adam Leive, Rolando Ossowski, Aaka Pande, Ajay Tandon, Jean Tesche, and Edit Velenyi. We also wish to thank Emiliana Gunawan for assisting in editing the document.

The author is also grateful to the World Bank for publishing this report as an HNP Discussion Paper.
I. INTRODUCTION AND KEY ISSUES

This paper summarizes the basic principles that should form the basis of fiscal policy. These principles encompass decisions on the functions of government, its spending, and the financing of its spending that affect economic growth, employment, inflation, and economic welfare. Although the principles are broadly applicable, it is especially important that health policy makers understand them. Ensuring access to health care is one of society’s — and therefore the government’s — most important goals. In meeting this goal, policy makers must be cognizant of fiscal realities — what they can expect the government to achieve in the health sector and at what cost? Resources are limited, and many valuable activities in both the private and public sectors vie for them. Allocating resources efficiently is of paramount importance, with implications for what the government does and how it finances its activities. The absolute level and share of government resources allocated to the health sector will depend on a variety of factors, but the bottom line is that health programs must compete with other government programs for scarce resources to ensure that these resources are put to their best use.

Recognizing these realities is increasingly important, as countries at all levels of development attempt to increase the quality of and broaden access to health care, with many countries setting a goal of universal coverage. Typically, these efforts require significant increases in government spending on health care, which raises the government’s share of total health spending. At the same time, the global economic crisis has increased the risk that donor aid to developing countries could stagnate or even decrease. One response has been for health policy makers to seek new sources of increased and dedicated funding. Although the motivation for this response is understandable, the problem of funding must be solved within the broader context of the overall budget. Communication between health policy makers and budget agencies must be improved so that all concerned parties can strive together for efficiency and equity in both government spending and revenue raising. As discussed below, earmarked funding for particular programs can often create more problems than it solves.

The overarching question to answer before a cohesive fiscal policy can be formulated is what the government role should be in society and, in particular, in the economy. Obviously, there is no single answer, and responses vary across countries depending on, among other things, social norms, levels of development, and administrative capacity. The answer to this question informs and provides the rationale for a myriad of choices about fiscal policy.

Once a choice is made about a government’s role in the economy, decisions can be made about which programs best fulfill that role — that is to say, a coherent fiscal policy can be formulated.

---

1. The term “sector” is used in two ways in this primer. On the one hand, it distinguishes the public from the private sector — one partition of the economy. On the other hand, it distinguishes different production and consumption sectors of the economy — for instance, transportation, health care, or national defense — another partition of the economy. Consequently, any economic entity will be in one sector of each partition — for instance, a private hospital is in both the private sector and the health care sector.
Expenditure policy is the core of fiscal policy. Milton Friedman said that the true level of taxation is equal to the level of spending. By this he meant that spending must be financed, and there are three ways to do so: (1) with current revenue; (2) by borrowing, which must in turn be repaid with future revenue; or (3) by printing money, which decreases the value of assets and flows that are fixed in nominal terms. Except for countries that rely heavily on grants or on exploiting natural resources, revenue is raised primarily through taxes, closing the causal loop between spending and taxation. Even when nontax revenue is available, marginal spending is financed almost entirely by current, future, or inflation taxes.

Two conditions should be met in allocating the budget across expenditure programs. First, resources must be allocated across programs in a manner that reflects their relative importance to society. Second, each program must be structured to maximize the value of the services that are either directly provided or financed. The focus of tax policy — and revenue policy more generally — is to raise the resources needed to run the government in as efficient and equitable a manner as possible. Unless a program is financed on a cost recovery basis — for example, by charging bridge tolls — revenue and expenditure policies are separable. The important connection is that the cost of raising revenue must be considered in deciding the scope of government spending. As will be discussed below, to maximize the benefit of government programs, the marginal cost of raising revenue must be less than or equal to the marginal benefit of government spending.

Section II presents issues to be addressed in determining the level of government spending and its distribution across programs. Section III identifies long-term fiscal challenges. Section IV presents the principles underlying the design and administration of revenue policies. Section V discusses the conditions to make fiscal policy sustainable. Section VI presents the implications of the issues as they relate to health care policy. The annexes offer more details on some of the concepts.

II. SETTING EXPENDITURE PRIORITIES

Perhaps a government’s most difficult task is to set spending priorities. Especially in a democratic environment, priority setting is the quintessential public choice problem. Politicians typically have a short time horizon — defined around the election cycle — so they hesitate to put the long-term interests of their constituents first. The problem is compounded by the fact that voters often reward political behavior that is counterproductive. It is difficult to align the public interest with politicians’ private interests when setting priorities (see box 1.1 for World Bank advice on this challenge). This section discusses the relatively easy conceptual aspects of setting priorities — addressing the twin goals of efficiency and equity. It offers several suggestions about how to address the public choice problem, but notes that, in fact, no easy solution exists.
BOX 1.1 WORLD BANK ADVICE ON ADDRESSING THE POLITICAL ECONOMY OF POLICY REFORM

Through the systematic combination of diagnostic and operational elements, the framework (developed in the book *The Political Economy of Policy Reform: Issues and Implications for Policy Dialogue and Development Operations*) aims to assist development practitioners to better understand and manage the political economy of reform, to design more equitable and sustainable reforms and operations, and thus to enhance development effectiveness.

In summary, the study points to the following three key messages for effective engagement by development agencies in processes of sector level policy reform and operations:

1. The importance of good political economy analysis, applied early in the process, to effective donor engagement in sector policy reforms and operations.
2. The significance of a sustainable process of building coalitions for change that involves dialogue with a broad range of stakeholders, including client governments, donors, other development partners, and the public.
3. The importance of promoting transformative processes of institutional change, including empowering forms of bottom-up accountability.

In terms of attitudes, behaviors and general professional practice within development agencies, the findings suggest that further emphasis should be given to

1. Getting a contextual understanding of the issues by combining analysis and dialogue, including insight into the likely distributional equity outcomes of any given reform, plus better access to reliable data which will improve operational design.
2. Broadening the range of interaction of development partners with a wider range of stakeholders (not limited to Finance or Sector Ministries), including sub-national or local governments; the parliament, including the political opposition; private sector; civil society, and formal and informal institutions.
3. Applying a partnership approach; based on listening and learning, more emphasis on participatory approaches for improved communication; and valuing and using local expertise wherever possible.

Taking a stronger process-oriented approach implies that practices and procedures of development agencies should become more flexible. In relation to the World Bank, the following areas emerged as important for enhanced agility and flexibility:

1. The choice and use of development support, including timing and sequencing of different lending and non-lending activities.
2. The design of lending operations, including timeframes and budgeting for loan preparation and implementation, and
3. Conducting and incorporating political (economy) analysis into the design, implementation and performance assessments of reforms and operations.


In theory, setting priorities is straightforward. The *marginal net social benefit (MNSB)* of the last dollar spent — that is, the benefit to society of the marginal government spending minus the cost of obtaining the resources for this spending — should be equal across all government programs and equal to the marginal net social benefit of spending on private programs. For instance, to finance an additional dollar of spending on health, the government must divert that dollar from some other use. If it diverts it from another government program, the benefit of spending the
dollar on health must be greater than from spending the dollar on the other program. Alternatively, if it diverts the dollar from the private sector — through taxation or borrowing — the benefit of the government health spending must be greater than the benefit that would accrue to leaving the dollar in the private sector plus the excess burden of raising the revenue (see below and annex 3).

This simple formulation — characterized by economists as allocative efficiency — hides intractable complexity. We must be able to measure net benefits — that is, benefits minus costs. Measuring cost accurately is difficult in itself. Measuring benefit is much more subjective, especially in areas such as health. No simple metric exists. Moreover, we must focus on social benefits and costs, which refer to the sum of private benefits and costs that accrue to individuals plus any additional benefit or cost that accrues to society as a whole. For instance, when consumers buy gasoline, they receive the private benefit of driving their cars. But there is also a social cost to society from the pollution emitted when the gasoline is consumed. Unfortunately, there is no unambiguous method for measuring social costs and benefits. We can define social externalities in concept — for instance with respect to pollution (negative) or education (positive), but measurement requires weighing the needs of one individual against the needs of another. Finally, we are concerned with the marginal benefit, again difficult to apply in practice.

A few simple equations can clarify these points within a static, one-period framework. Let

\[ MSB = MPB + MXB, \]

where \( MSB \) equals marginal social benefit of spending on a particular activity — say a government program or the consumption of a good or service, \( MPB \) equals the marginal private benefit individuals derive from the activity, \( MXB \) is marginal benefit society derives from positive externalities associated with the activity. The marginal social benefit must be at least as great as the marginal social cost of the last dollar spent on the activity (\( MSC \)):

\[ MSC = MPC + MXC + MDWL, \]

where \( MPC \) equals the marginal private cost incurred by individuals on the activity (including any taxes imposed on them), \( MXC \) is the marginal cost society derives from negative externalities associated with the activity, and \( MDWL \) is the marginal deadweight loss/excess burden a government imposes on society to raise the resources needed to finance its activities.

Combining equations (1.1) and (1.2), the efficiency condition is \( MSB \geq MSC \). For market-based activities with no externalities, the efficiency condition reduces to \( MPB \geq MPC \). Where externalities exist, the government must intervene to insure both positive and negative externalities are addressed with, for example, subsidies or taxes. When government funds are involved — say, when the government finances the activity or subsidizes a private market — efficiency must also account for the deadweight loss incurred in raising the resources the government uses.

Although it is not possible to apply this conceptual framework in practice, it does provide a frame of reference for setting priorities and pursuing allocative efficiency (see annex 1 for more
detail on allocative and productive efficiency). Using judgment rather than precise measurement, a simpler — but still difficult — approach would entail the following steps:

- Identify the appropriate scope of government — that is, which functions should lie in the purview of the government and which should be left to the private sector.
- Focus on the functions and programs undertaken by government — for instance, national defense or education, rather than an economic classification of spending — such as wages.
- Make informed judgments that reflect economic and political realities.
- Learn as much as possible from international experience and avoid pitfalls.

2.1. THE PROPER SCOPE OF GOVERNMENT

Opinions about the activities for which the government should take direct responsibility vary widely (see annex 2). It is broadly agreed that governments should provide pure public goods, such as national defense or police protection, (a pure public good is one that is both nonexcludable in that an individual cannot effectively be excluded from consuming it and nonrivalrous in that consumption by one individual does not reduce the availability of the good to others). Less agreement exists about the level of government involvement for other services, ranging from health and education to electricity generation and trash collection. Although a government may have a legitimate role in ensuring broad access to these services, this goal can be achieved in a variety of ways short of direct government provision. In making an informed judgment on where public provision — either through direct production or financing — stops and private provision begins, it is important to be clear about the criteria for and implications of a particular choice.

The first question to be answered is whether market forces can be brought to bear in rationing a particular service, whether it is publically or privately provided. By definition, the marginal cost of providing a pure public good, such as defense, to an additional recipient is zero. Market mechanisms rely on setting price equal to marginal cost, so the market has little to offer in rationing national defense. Rather, public goods are provided on a tax and transfer basis: the government collects taxes to finance national defense, which is then provided at no additional cost to the citizens. Providing government services on a tax and transfer basis has several implications:

- First, the quantity provided to each recipient cannot be rationed by price. This has no implication for a pure public good, for which consumption by one person has no effect on the cost of providing the service or the amount of the good available for others. However, it does have implications for other types of goods that are often provided by governments, such as basic health services, roads, or national parks.

- Second, if the supply of a particular service is not rationed by price, the incentive for consumers to limit the amount they consume must be provided in another manner. For example, the incentive to limit road use can sometimes be congestion, limiting a consumer’s
desire to use the road. For a normal good — one with a flat or upward-sloping supply curve and a downward-sloping demand curve — nonprice rationing is inefficient.

- Third, the provision of the service, including through private contractors, must be financed through taxation, borrowing, or printing money. Each has inherent costs. Except in special circumstances, taxation distorts economic activity. If the government borrows, the tax is deferred to future generations. Printing money causes more rapid inflation, which is another form of tax. Conversely, a tax that is not spent reduces the need for future taxes, so it simply changes the timing of taxation, not adding to its level.

The provision of goods on a tax and transfer basis is more costly than when provision is price rationed because virtually all taxes distort efficient production and consumption choices, whereas competitively determined market prices maximize efficiency (Annex 3). Moreover, it is costly to raise revenue. As discussed above, a tax will almost always engender excess burden or, equivalently, deadweight loss. Other than taxes that have either no or a beneficial impact on behavior — for instance, a lump-sum (poll) tax, a tax on a good for which demand is completely inelastic, or a tax on a good that has a negative externality, for example, carbon fuels) — the deadweight loss of taxation can add significantly to the cost of government spending.

In addition, taxes — both the structure of the tax system and the size of revenues — can affect individuals’ willingness to save and invest, thereby affecting economic growth. Substantial research has been done on the implications of the size of government on economic growth (for instance, Alesina and Perotti 1996). Although the results are not definitive, it appears that beyond a point, additional government spending can retard growth.

Various mechanisms avoid or mitigate the deleterious effects of tax and transfer schemes:

- For those programs for which tax and transfer mechanisms are absolutely necessary — for instance, the provision of pure public goods or redistributive spending — the goal should be to reduce deadweight loss as much as possible and limit the impact of the tax system on incentives to save and invest. For public goods, this requires making informed judgments about the optimal supply of a public good and maximizing productive efficiency. For redistribution programs, this requires good program design and effective targeting — limiting errors of inclusion and exclusion — to maximize the social benefit of redistribution and minimize the cost of financing.

- For goods or services that do admit to price rationing, the government can (1) produce the good or service, (2) finance production by contracting with the private sector, or (3) regulate the private sector. The more developed the governance of an economy — through, for example, clear property rights and effective contract laws — the less government must rely on itself to produce goods and services, and the more it can focus on financing and regulation (Shleifer 1998). In some cases, a mandate to households or employers to purchase a good or service, combined with transfers to poor and vulnerable households to ensure they can afford the direct cost or the compensating reduction in their wages, can achieve a social objective more efficiently (Summers 1989).

- Where equity is an issue, the government can raise resources to finance its activities in a progressive manner and substitute well-designed resource transfers that allow poor and vulnerable households to obtain needed goods and services without having to interfere with
market mechanisms. With respect to the latter, the work of the World Bank on targeted transfer mechanisms, including conditional cash transfers, presents criteria and methods for implementing such programs (Grosh et al. 2008; Fiszbein and Schady 2009). This work demonstrates that targeted transfers can operate and be effective even in developing countries.

- The same choices on how governments should support or even supplant private markets apply to the health sector. Even if a society decides the government should be the single payer for health care, it can rely on the private sector to provide the services, using prices to help ration services, recognizing that several pricing models — for instance, fee for service or capitation — are available. Government financing and regulation can help address problems of equity, asymmetric information, consumption externalities, moral hazard, and adverse selection. Finally, government can address equity goals by focusing its market interventions to help poor and vulnerable households. Market failures are not qualitatively different in the health sector, although they may have greater quantitative impact. What this implies for the government role is subject to debate.

2.2. FOCUS ON GOVERNMENT FUNCTIONS AND PROGRAMS

The size of government varies widely across levels of development. The median value of general government spending as a share of GDP was 32 percent for 2010. However, the median shared ranged from 24 percent in low-income countries up to 44 percent in high-income countries (figure 1.1). “Wagner’s Law” (posited in 1863 by Adolph Wagner) provides a stylized rationale for this empirical observation. “The law is based on three claims: economic growth results in an increase in complexity requiring continued introduction of new laws and development of the legal structure; urbanization increases negative externalities, such as congestion and crime, which necessitate intervention; and the goods supplied by the public sector have a high income elasticity of demand. If the elasticity of demand exceeds one, public sector expenditure will consequently rise as a proportion of income” (Black, Hashimzade and Myles (2009), p. 481). Whether Wagner’s Law on government behavior holds in general is an open question, but the correlation between the size of government and economic development is evident. In some sense, large government is a “luxury” that is — at least on average — more affordable at higher levels of income.

The size of government also varies significantly within development categories, belying a strict adherence to Wagner’s Law. For instance, within high-income countries, expenditure as a share of GDP ranged from 15 percent in Singapore to 65 percent in Ireland, with a standard deviation of 11 percentage points. The standard deviation within low-income countries was 7 percentage points, with the share ranging from 12 percent in Myanmar to 41 percent in Burundi. The variance within income groups far outweighs the variance among different income groups. Finally, median spending increased within each income class between 2000 and 2010. This reflects the natural tendency of the size of government to grow with income level, as well as the impact of the global recession on both spending (positive) and GDP (negative) in 2010. It is interesting that the IMF projects the median ratio of expenditure to GDP to fall in all but the low-income category between 2010 and 2015.
There is no analytical solution for the optimal size of government. The choice reflects societal values as well as the economic issues raised above. There is no doubt inefficiency in government production of goods and services — productive inefficiency — as there is also in the private sector. Reducing inefficiency in the government’s direct production of goods and services or in private sector production that the government finances would allow the same goods and services to be provided at lower cost, an obvious improvement. However, reducing “waste, fraud, and abuse” — as productive inefficiency is often characterized — does not address the fundamental question about the appropriate role of government.

To improve both allocative and productive efficiency, it is important to focus on the functions and programs of the government (table 1.1). The functions across the top of table 1.1 describe what the government produces — its outputs. The economic classification, on the other hand, focuses on inputs — what the government consumes in the production process. Both are important, but in evaluating the appropriate role of government, outputs are far more important. (One could go a step farther and focus on outcomes — for instance, an outcome in the health sector might be the increase in life expectancy that is achieved with new treatments, rather than the treatments themselves. A focus on outcomes comes very close to looking at marginal net social benefit. Outcomes are a useful classification for informing the policy debate, but not operational on a day-to-day basis.)

To produce output, the government should combine its inputs efficiently — productive efficiency — producing a particular set of outputs at their lowest cost. The more difficult question is which outputs should the government produce — allocative efficiency. To do this, we must look within the functional classification at the individual government programs. Moreover, we have to carefully evaluate all off-budget government activity in, say, state-owned enterprises and public-private partnerships. The point is not to eliminate the activity, but to conduct it
efficiently, equitably, and transparently. It is important to emphasize that the government must compete not only with the private sector for resources to improve allocative efficiency, but each government program must compete with every other government program to improve allocative efficiency within the public sector.

### Table 1.1 Alternative Expenditure Classification Schemes

<table>
<thead>
<tr>
<th>Economic classification</th>
<th>Functional classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Health</td>
</tr>
<tr>
<td>Current</td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td></td>
</tr>
<tr>
<td>Good &amp; services</td>
<td></td>
</tr>
<tr>
<td>Interest</td>
<td></td>
</tr>
<tr>
<td>Subsidies &amp; transfers</td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td></td>
</tr>
<tr>
<td>Net lending</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Authors*

#### 2.3. Make Informed Judgments

Expenditure policy is at least as much art as science. Information and data are crucial for making good decisions, but so is good judgment. Setting long-term priorities is, by definition, a long-term exercise, and a government cannot be restructured instantly. The best approach is likely to be incremental, devising a transition strategy toward the desired structure of government that allows for continuous review and adjustment. It is important, however, that the incremental steps — whether to restructure government or address shorter-term fiscal problems — be consistent with a long-term strategy encompassing both policy goals and fiscal sustainability.

Although it is impossible to do major restructuring quickly, there are several rules of thumb with respect to spending that can be useful in the short term:

- Wage spending can often crowd out spending on other inputs. Because wages are viewed as nondiscretionary spending, they often take precedence over other inputs. In the health sector, spending on salaries might crowd out spending on pharmaceuticals, disposables, and maintenance. In education, it might crowd out expenditure for books or school uniforms.

- The mix of spending in health and education is often flawed. Too many resources are spent on tertiary health care and higher education and not enough on primary care and primary and secondary education, even though the social rate of return to primary education and primary care are higher (see Birdsall 1996 for a discussion of rates of return to spending on education). The reforms require improvements in both allocative and productive efficiency.

- Even more difficult, the mix of spending across programs can become unbalanced. As a society ages, for example, resources should likely be shifted out of education into health care and old-age income support.
Price subsidies are almost always inefficient. To the extent they are justified to protect vulnerable households, they are often too expensive and badly targeted; much better mechanisms exist to achieve the same objective.

Government investment is potentially very productive, but it is also seductive for politicians, who can reap political rewards. Measuring even the private rate of return on an investment is difficult; measuring the social rate of return is more problematic. It is important to closely scrutinize investment projects and include all future operations and maintenance costs in balancing costs and benefits.

The incidence of spending — who actually benefits — should be carefully analyzed. Benefits can often accrue disproportionately to higher-income households (see Gillingham et al. 2008, for an analysis of distributional incidence). Since these same households either pay or should pay a disproportionate share of taxes, the government ultimately creates deadweight loss to collect revenues and then distributes resources back to the same households that paid the taxes. One needs to assess both the benefit and revenue incidence as well as the net (overall) incidence.

It is important to start any needed reforms as early as possible, based on a well-sequenced transition plan. As the next section notes, the future is fraught with serious fiscal risk. Reforms in programs such as pensions and health are politically difficult and must be carefully phased. It is costly to wait.

Across-the-board spending cuts are often criticized as inefficient. However, if spending is allocated efficiently to start — that is, marginal net benefits are equal across programs — an across-the-board cut is optimal, at least at the margin. It is only when spending is inefficiently allocated — as is common — that individual programs have to be targeted.

### 2.4. AVOID PITFALLS

When budget pressures are severe, it is natural to look for ways to pursue government objectives that do not add to these pressures. Examples include creating off-budget entities, such as state-owned enterprises to perform particular functions or to pursue public-private partnerships that can obscure public costs. In both examples, the program typically produces an output that is reasonably well-defined and can be a candidate for privatization. Where functions are not shifted to the private sector, long-run costs and fiscal risks must be transparent. An equally problematic example of poor governance is to substitute regulation for direct government action. This may reduce the apparent “size” of government, but at the cost of serious unintended consequences. For instance, a government may mandate employers to provide health insurance for their workers, without adequately recognizing the impact of the mandate on wages and on the efficiency of the labor market.

To help avoid pitfalls, a government can establish budget procedures that ensure transparency, careful review, and informed judgment. The medium- and long-term implications of budget decisions need to be estimated. The likely economic and social impact of important programs should also be analyzed, with independent, external input where possible. Off-budget entities should be avoided absent a compelling public purpose. They lack transparency and increase the
risk of unforeseen fiscal costs if and when these entities fail. Fiscal rules — laws or regulations that limit some combination of spending, revenue, and the deficit — can be considered, but any fiscal rule must be carefully structured to avoid being overly restrictive or slack. The purpose is to reduce fiscal flexibility; however, the loss of flexibility can constrain governments from taking actions that are in the public interest. Anderson and Minarik (2006) make a strong argument for focusing fiscal rules on expenditure levels, rather than on the deficit.

III. LONG-TERM EXPENDITURE CHALLENGES

Two spending categories that promise to severely strain government budgets in the future are pensions and health care. Moreover, in many countries, these two sectors compete for revenues from the same earmarked payroll levies. Regarding pensions, pressure will arise due mainly to the aging of the population, which is already well advanced in more developed countries and will increase rapidly in less developed countries in the future (figure 1.1). As this process evolves, lacking reform, the number of persons collecting benefits relative to the number contributing to a scheme will increase significantly. Cash-flow deficits in these schemes already exist in many countries and are projected to begin in the near future in many others. At that time, governments will be called upon either to honor securities they issued to pension trust funds or finance the cash flows directly out of current revenues or borrowing.

Health care expenditures will also grow with the proportion of elderly in the population, but they are more affected by demand factors and the rapid growth in health care costs. Both of these factors are driven by rapid technological change and increasing incomes. Although the former reduces prices — for example, the price of cataract surgery has fallen while at the same time the quality of the surgery has increased (Shapiro, Shapiro, and Wilcox 2001) — it also expands the options for treatment, leading to an increase in the quantity of health care services consumed. Even adjusted for population aging, health care spending has generally grown faster than GDP as a result of these innovations. The problem is acute in many developed countries. In less developed countries, where the quantity of services consumed is low, both total and government spending are similarly low. Thus, the challenge in these countries will be to allow the population to enjoy at least some of the fruits of modern medicine without creating unsustainable fiscal pressure. In other words, although it is seductive to argue for greater health spending, the increase in spending, as well as any improvement in health outcomes it produces, must still be justified on the grounds that it yields the highest marginal net social benefit. Advocates for all government programs must recognize they are competing for resources, and the efficacy of spending should be the metric for how to allocate resources.

2. Another argument for health spending is that it increases the capacity of people to work, which, in turn, increases labor productivity and output. This is a valid argument for health spending — although the effect should not be overestimated — but it is not necessarily an argument for government health spending. The benefits of greater productivity can be captured by workers and used to finance health care, except, perhaps, for the poor and otherwise vulnerable, who are not in a position to make this investment in human capital.
3.1. Public Pensions

Public pension systems take two primary forms. First, broad-based schemes are designed to cover all or almost all workers, although they may only reach workers in the formal sector. Second, some countries have separate pension schemes for civil servants, military personnel, and other special categories. Most pension systems are at financial risk and represent a large and growing burden on central government finances. Many public pension systems did accumulate large reserves when they were first implemented, a large percentage of which were — or are — in the form of government-issued securities. As these schemes matured, their cash flows fell, and many began to run cash-flow deficits; the result has been growing pressure on government finances as government securities are liquidated. Absent reform, the fiscal pressures will spread to more countries. Moreover, if reforms are delayed, adjustments will have to be larger and more abrupt, and more of the reform burden will shift to younger generations.

Civil service schemes are typically run on a pure pay-as-you-go basis. Benefits are paid directly from current revenues, some of which may be from pension contributions, or the accumulation of new debt. Again, lacking reform, cash-flow deficits will increase as growth in the number of covered employees slows, the number of retirees grows, and beneficiaries live longer.

Although a serious policy challenge, designing a pension system that is fiscally sustainable is analytically straightforward. At least prospectively, contributions — plus any transfers from the budget to achieve equity goals — must be sufficient to finance benefits. This goal can be achieved in a number of ways, including adjusting the parameters of a defined-benefit pension system or transitioning to defined-contribution or notional defined-contribution system (Gillingham 2003. Once prospective balance is achieved, the question is how to finance any unfunded liabilities accrued up to that point. This, again, is a policy decision concerning the
extent to which existing pensioners and workers near retirement should share in either reducing unfunded liabilities or in financing them.

3.2. HEALTH CARE

The share of GDP devoted to health care increases with the level of development (figure 1.3). Given the variation in incomes across levels of development, the absolute amount of health spending varied in 2010 from an average of US$41 in low-income countries to an average of US$2,989 in high-income ones (figure 1.4). Low-income countries spend a significant portion of GDP on health care — in fact, close to the share in middle-income countries. Unfortunately, GDP per capita is so low that the absolute level of per capita spending is barely one-quarter of the level in lower-middle-income countries and less than one-thirtieth of the level in high-income ones (figure 1.4).

Spending increased significantly, both in absolute terms and as a share of GDP between 1995 and 2010 (see Fleisher et al. 2012 for a discussion of trends). There are many reasons for this (CBO 2008). Three factors that are essentially universal are the aging of populations, increases in real income, and technological change. First, the elderly typically spend more per capita on health care than those who are younger, so the increase in old-age dependency ratios has resulted in an increase in the demand for health care. Cutler (2003) assumes that the elderly (65+) spend three times as much per capita on health care as those age 15 to 64, and those under 15 spend 60 percent less than the adult population. This has had a higher impact on spending in higher-income countries, where aging is more pronounced; however, its impact until now has been relatively low, explaining only about 2 to 3 percent of the increase in real spending on health care in the United States, where increases in health spending have been the most rapid (CBO 2008). Second, most estimates of the income elasticity of demand for health care are greater than one. If this is true, the share of health care in total consumption will increase as incomes increase.

Third, the rate of technological change in health care has been very rapid. Treatments — both preventive and curative — that were impossible just a few years or decades ago are now...
common, at least in high- and middle-income countries. Technological change has been broadly based, improving available drugs, medical devices, and clinical procedures. Again for the United States, the CBO (2008) and others (including Newhouse 1992; Cutler 1995; and Smith, Heffler, and Freeland 2000) estimate that over half and as much as two-thirds of the increase in real health spending can be explained by technological change. The paradox is that technological change reduces the price of health care, but expands the available treatment options by so much that it increases the level of spending on health care. One challenge going forward will be to allow the poor, especially in poor countries, to reap at least some of the benefits of this new technology.

Government spending on health care has grown pari passu with total health spending (figure 1.5). The slight upward trend in the average share of government spending across all countries reflects the fact that the share increases with the level of development. As income grows, so does the role of the government as, at least, a source of financing for health care. This, in turn, increases the share of total government spending allocated to health care (figure 1.6), and so imposes severe pressure on government budgets in developed countries. As less developed countries advance, these pressures will be felt more broadly. They will witness at the same time other long-term fiscal challenges — including not only pensions, but, for example, any costs to mitigate the impact of climate change (see Heller 2003 for a detailed discussion of long-term fiscal challenges). Thus, the importance of setting priorities will only increase, which is why governments must consider if at least some public services could be rationed by price. In this case, even though the government is the financer or direct supplier, market forces are brought to bear and the deadweight loss of tax financing is curtailed.

**Figure 1.5 Government Health Spending as a Percent of Total Health Spending**
(Unweighted average)

**Figure 1.6 Government Health Spending as a Percent of Total Government Spending**
(Unweighted average)

Source: Leive et al. 2012.

**IV. RAISING REVENUE**

Taxes and other forms of revenue, such as from the sale of natural resources, donor aid, and user charges, play various roles in fiscal policy, but first and foremost is raising revenue to finance government spending. Oliver Wendell Holmes, Jr., a noted American judge, once said “Taxes are the price we pay for a civilized society” (quoted by the US Internal Revenue Service above
the door of its headquarters in Washington, DC). By this he meant that the government needs resources to fulfill its essential functions, and someone must pay the bill. Taxes and other revenues might also be used to (1) redistribute income and wealth, and (2) shape economic behavior. For all these purposes, however, a necessary condition to justify a tax is that it raises revenue.

Revenue levels vary widely across countries. As with expenditure, the median value of revenue as a share of GDP increases with a country’s level of development (figure 1.7). In 2010, the median ratio of revenue to GDP ranged from 21 percent in low-income countries to 39 percent in high-income ones. Thus, an average low-income country, with a 2010 per capita GDP of $550, can raise $120 per capita in government revenues, while an average high-income country, with a per capita GDP of $34,573, can raise $13,873 per capita in government revenues. It is not surprising that ratios of revenue to GDP roughly mirror ratios of spending to GDP. As noted earlier, in the medium to long term, the level of spending is arguably the true level of taxation in an economy, and collecting the taxes concurrently with spending simplifies government finance and avoids the accrual of liabilities that must be honored later (or, of course, defaulted). The next section discusses conditions for fiscal sustainability. Median revenue-to-GDP ratios increased for countries in all income classes from 2000 to 2005. Except for low-income countries, the growth was slowed or reversed between 2005 and 2010, largely due to the economic crisis. As with spending, the variation in revenue-to-GDP ratios within income classes far exceeds the variation across them.

**FIGURE 1.7. GENERAL GOVERNMENT GROSS REVENUE**

(Group median as a share of GDP)

<table>
<thead>
<tr>
<th>Year</th>
<th>Low income</th>
<th>Lower-middle income</th>
<th>Middle income</th>
<th>High income</th>
<th>Upper-middle income</th>
<th>All countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>10.5%</td>
<td>14.5%</td>
<td>20.5%</td>
<td>26.5%</td>
<td>32.5%</td>
<td>20.5%</td>
</tr>
<tr>
<td>2000</td>
<td>11.0%</td>
<td>15.0%</td>
<td>21.0%</td>
<td>27.0%</td>
<td>33.0%</td>
<td>21.0%</td>
</tr>
<tr>
<td>2005</td>
<td>11.5%</td>
<td>15.5%</td>
<td>21.5%</td>
<td>27.5%</td>
<td>33.5%</td>
<td>22.0%</td>
</tr>
<tr>
<td>2010</td>
<td>12.0%</td>
<td>16.0%</td>
<td>22.0%</td>
<td>28.0%</td>
<td>34.0%</td>
<td>22.5%</td>
</tr>
<tr>
<td>2015</td>
<td>12.5%</td>
<td>16.5%</td>
<td>22.5%</td>
<td>28.5%</td>
<td>34.5%</td>
<td>23.0%</td>
</tr>
</tbody>
</table>

*Source: IMF, World Economic Outlook, October 2012 database.*

There has been broad consensus, since as early as Adam Smith in 1776 and continuing to the present (see, for instance, Heady 1993, on the traits of a good tax system), that a revenue system should do the following:

- Promote growth and efficiency
Be equitable
Be easy to administer
Be understandable and predictable (transparent)

The rest of this section will summarize ways to raise revenue that are consistent with these criteria.

For most countries, tax revenue is the primary source of government financing. The goal of tax policy is to raise sufficient revenue without unduly distorting the economic behavior of consumers and producers, while still allocating the tax burden equitably. The key policy challenge is to choose a tax base — that is, an economic flow or stock — on which to impose the tax. The three most popular choices are consumption, income, and wealth. Bradford (1986) provides an intuitive framework for discussing taxation by focusing on income and then framing the policy choice as determining the portion of income that should be taxed. He starts with “Haig-Simons” income (named for the two economists who developed the concept), defined as consumption plus net saving. This definition focuses on the use of income. Income can also be defined in terms of its sources, starting with labor income and returns to saving (factor incomes) and subtracting net saving (for instance, deposits to savings accounts less withdrawals). The focus on income, including on its sources and uses, clarifies the distinctions among many of the taxes used today. The rest of this section uses this framework to discuss three of the most commonly used bases for taxation — consumption, income, and wealth — and then the sources of and issues related to nontax revenues. (See box 1.2 for a summary of IMF advice on reforming tax systems in developing countries.)

4.1. CONSUMPTION TAXES

Consumption Taxes

A broad-based consumption tax — which focuses on the use of income for consumption — is an effective way to collect tax revenue. From an efficiency perspective, it exempts the rate of return to saving from taxation. This, in turn, yields “temporal neutrality” — that is, it does not distort the relative prices of current and future consumption, and therefore does not distort the intertemporal allocation of resources. Consequently, the (discounted value of the) taxes imposed on two taxpayers who differ only in the share of the income they save — that is, how much consumption they defer — will be the same. Second, from a practical perspective, a broad-based consumption tax focuses collection on a relatively small number of taxpayers.

---

3. The literature on tax policy is extremely broad and diverse. This section can only touch on some of the most important elements of that literature. Examples of detailed and practical expositions of tax policy include Bradford (1986), Mirrlees et al. (2010), and Mirrlees et al. (2012). (The Meade Committee Report [Institute for Fiscal Studies 1978] was an important precursor to these references.) Bradford focuses primarily on the United States and the two Mirrlees volumes, on the United Kingdom, but all treat tax policy issues in a manner that is broadly useful.
BOX 1.2 COMMON ELEMENTS OF STRATEGIES FOR REFORM

IMF advice to developing countries has commonly stressed:

- Establishing effective revenue administrations making proper use of withholding and third-party information, and capable of building on these to implement voluntary compliance and self-assessment — taxpayers calculating and remitting tax themselves, subject to audit and penalties — both as a prerequisite for expanding the tax base and addressing corruption.

- Assuring strong control of the largest taxpayers, in a dedicated office (and with specialized units for the most critical sectors), as a key step toward introducing risk assessment and fuller taxpayer segmentation.

- Implementing policies and procedures that limit opportunities for rent seeking and help identify and punish inappropriate behavior in the revenue administration.

- Designing and applying forceful and efficient strategies to deal with noncompliance.

- Ensuring that laws and regulations are reasonably simple, readily available, coherent across taxes, and provide good taxpayer protection (including effective appeals procedures).

- Replacing inefficient production or sales taxes, after adequate preparation of both the administration and taxpayers, by a simple VAT — including catalyzing administrative reforms.

- Levying a VAT on a broad base, with a high threshold (the level of turnover at which registering for the tax becomes compulsory) and avoiding multiple rates, to realize its potential as a reasonably efficient source of government finance.

- Coordinating any prospective loss of trade tax revenue with measures to replace it from domestic sources.

- Avoiding exemptions — under all taxes — that jeopardize revenue and good governance, are hard to reverse, and generate no clearly offsetting social benefit.

- Removing minor taxes and fees that are inordinately costly to comply with and administer.

- Building corporate income taxes that are simple (in their depreciation and carry forward provisions, for instance) and sufficiently broad-based to allow statutory rates competitive by international standards, with effective tax rates that are reasonably low and uniform across investments.

- Strengthening capacity to deal with profit-shifting by multinationals, while recognizing the extreme difficulty of doing so.

- Extending coverage of the personal income tax (particularly through inclusion of smaller businesses and professionals) and establishing coherent taxation of capital income, with an effective rate structure consistent with the authorities’ distributional preferences.

- Exploiting the potential for regional cooperation, in both policy and administration — particularly on business taxation and excises — to limit mutually damaging competition.

- Balancing royalties, auctioning, and profit-related charges in taxing natural resources.


The value-added tax (VAT) — the most popular variant of a broad-based consumption tax — has spread through the world like wildfire over the past three decades (see Ebrill et al. [2001] for a comprehensive description of the VAT, and Crawford, Keen, and Smith [2010] for a more recent discussion of the issues surrounding indirect taxation). In addition to the traits of any broad-based consumption tax, it promotes efficiency for the following reasons:

- It is collected on all sales of goods and services, obviating the identification of consumption.
➢ It does not “cascade,” since the tax collected at intermediate stages of processing is rebated.
➢ It has the potential for higher revenue because it captures economic activity throughout the production and distribution chain.

Trade flows are unaffected because exports are zero-rated. However, as with any consumption tax, the VAT will reduce the incentive to work, since it does not apply to the consumption of leisure. A consumption tax that is inversely proportional to the elasticity of demand — the Ramsey rule — is most efficient (minimizes excess burden). But, this type of precision is impractical and can conflict with equity goals, since the demand of low-income households for necessities could be very inelastic. Hence, broaden the base and lower the rate. Multiple rates, exemptions, and zero-rating (except for exports) are inefficient. Moreover, they create serious problems for tax administration.

**Base Broadening**

Differential rates (with exemptions for specific goods and services as a special case) curtail the revenue generated by broad-based consumption taxes. Equity grounds are often used to justify differential rates, in particular for food. However, the concomitant tax subsidy is badly targeted. Although poorer households spend a larger share of their budgets on food, better-off households spend much more on food in absolute terms and, consequently, receive an arguably unfair share of the tax subsidy. Where possible, other redistributive mechanisms that better target the poor and vulnerable should be used to more effectively meet equity goals. The necessary condition for adopting differential rates is well stated in Mirrlees et al. (2012, p. 159):

> To distinguish specific egalitarianism (that is, differential rates) from the argument for generalized redistribution, we must believe that people will still choose to buy ‘too little’ of these goods even if they have the money to do so. The goal is to encourage people to buy these goods in particular rather than giving them enough money in general. In one sense, this is a more coherent argument for differentiation than is the general equity argument, since we could not achieve this outcome more efficiently using other tools that we currently have. On the other hand, it is an argument that sits uncomfortably with a belief that people are generally able to make the right decisions for themselves.

Equity grounds are also used to justify the standard exemptions from consumption taxes for health and education. For these two consumption categories, it is also often argued that (1) their consumption yields positive externalities and therefore deserves tax subsidies; and (2), to the extent they are provided on a tax and transfer basis, it would be difficult to capture their consumption in the tax base. All these rationales warrant careful examination before they are accepted. With respect to equity, the tax preferences for health and education share the defects of differential rates generally, with the benefits accruing disproportionately to richer households. The externality justification is strongest for certain components of health and education, for instance, for spending on immunization and public health or on primary education. It is arguable whether all spending on health and education deserves to be subsidized on efficiency grounds.

The argument for excluding the consumption of owner-occupied housing and financial services from the VAT rests primarily on the fact that the tax on these categories of consumption would be difficult to implement. These exemptions, however, significantly narrow the tax base, increase
the rate necessary to raise a target level of revenue (by more than the decrease in the base, since the higher rate elicits greater evasion and avoidance), and distort consumption choices. The IMF, in Claessens et al. (2010), discusses a method for bringing financial services into the VAT tax base.

Applying a consumption tax to government services provided as in-kind transfers is also difficult in practice, but exempting them has unintended consequences, especially for a VAT. Since the exemption applies only to the value added produced by the government supplier, the VAT collected on material and service inputs would still be collected. Unless the VAT paid on inputs flowed back into the budget of the agency making the choice to outsource, this exemption creates a disincentive for the government to contract out with the private sector, even if it would be more efficient. As discussed above, cash transfers, with which households can buy goods and services from either private or public providers, are an alternative to in-kind transfers that does not suffer from this problem.

Finally, a significant share of consumption escapes taxation either because (1) the services are not explicitly traded, or (2), especially in developing countries, the trades take place in informal markets. The primary example of the first category is owner-occupied housing. In the second category, an establishment must typically exceed a threshold size before it is required to collect VAT on its sales. Keen (2012) argues that economic agents operating in the informal sector are not an important potential source of tax revenue, so exempting small enterprises is relatively innocuous.

VAT As a Replacement for Trade Taxes

Trade taxes are a minor source of revenue in developed countries. In developing countries, however, trade taxes — primarily import tariffs and, to a lesser extent, export tariffs — have been an important revenue source. This is primarily because they are relatively easy to levy and collect, and import tariffs can protect domestic producers. As the world economy has become more open, the importance of trade taxes has decreased. Consequently, LICs have had to find alternative revenue sources, with mixed success. One benefit of the VAT is that it is also applied to imports (although refunded on exports) and is thus a way to raise revenue without putting LICs at a competitive disadvantage in world trade. Equally important, the VAT is consistent with broad-based consumption taxation, whereas tariffs apply indiscriminately and can distort production. Emran and Stiglitz (2005) take a different position on the efficacy of the VAT in LICs, focusing on its implications for the informal sector. Keen (2012, on the other hand, argues that the VAT, carefully implemented, can effectively replace the revenue lost to trade liberalization.

Excise Taxes

In addition to broad-based taxes such as the VAT, consumption can be taxed more selectively to both enhance revenue and improve efficiency. For simplicity, these selective taxes can be combined under the heading of excise taxes that can address “externalities” — that is, costs or benefits that accrue to society as a whole rather than to private participants in the exchange of consumption goods. This type of excise tax is often denoted as a Pigovian tax, after the economist who defined their conceptual framework. Such taxes have become more popular in
recent decades in addressing environmental issues. If the tax is calibrated to the social harm of consuming a particular good or service, it internalizes the full social cost of consumption on the participants involved in the exchange. To the extent it meets this goal, it can improve efficiency. The appropriate level of a Pigovian tax is difficult to ascertain (see, for instance, the discussion of gasoline externalities in Parry and Small [2005]), and is sensitive to interactions with other taxes (Goulder 1995).

A related set of excise taxes focuses on *bads*, such as tobacco and alcohol. Some level of excise tax on these commodities can be justified by externalities, but the levels of taxation are often much higher. Moreover, the incidence of the tax is highly regressive. Another rationale for these taxes is that consumption of these goods is addictive and consumers’ behavior toward them is “time-inconsistent.” That is, consumers — especially young ones — are unable to control their current behavior in response to the future implications of developing an addiction. This behavior creates what Gruber and Köszegi (2001) call an “internality” and Viscusi (1995) calls “an externality to one’s future self.”

In the face of such inability to control their behavior, some consumers may prefer higher taxes to increase the disincentive to consume addictive substances. However, the preferred tax is likely to vary widely among those who smoke and drink (in fact, moderate consumption of alcohol may yield long-term benefits), and the tax would also penalize those consumers who choose to smoke and drink even though they are aware of the long-term implications (these latter consumers act in a manner consistent with the “rational addiction” model of Becker and Murphy [1988]). With either intra- or international variation in the level of taxes, a high tax can also provide an incentive for smuggling. Finally, the concept of an internality also begs questions about differential taxes on a wide variety of goods that can be construed as having longer-term implications that consumers do not fully internalize.

### 4.2. INCOME TAXES

**Personal Income Taxes**

*A consumption-based personal income tax.* A direct tax on income can be used to capture the same tax base as an indirect tax on consumption. Most closely related to a VAT would be a two-tiered cash-flow tax that would tax value added less payments to employees at the business level and all compensation for labor at the personal level (see Hall and Rabushka (1985), for example). In practice, with labor costs accounting for the largest share of value added, the bulk of tax revenue under the two-tier cash-flow tax would be collected at the personal level. With a single flat rate at both the business and personal levels, this tax would be identical to the VAT. It would simply shift labor compensation from the business tax base to the personal tax base and would offer the option of taxing labor compensation progressively — that is, with an average tax rate that increases with the level of compensation. On the other hand, it greatly expands the number of taxpayers, with concomitant administrative complexity. This presents a particularly daunting challenge in developing countries.

Just as with a VAT, the breadth of the tax base under the two-tiered cash-flow tax is a policy choice. With only a single, fixed exemption per taxpayer, the personal cash-flow tax would allow
no special deductions for mortgage interest, state and local taxes, financial services, or spending on health and education. Since private-sector producers of these goods and services would not be exempt from the business cash-flow tax, and their labor payments would be treated the same as any other labor compensation at the personal level, these spending categories would largely fall into the tax base. The value added of goods and services provided as in-kind transfers from the government would be exempt from business-level taxes, creating the same qualitative disincentive to contract out to the private sector, but, because the value-added base is much narrower, the disincentive would be smaller. As with indirect consumption taxes, the goal should be to broaden the base and lower the rate to mitigate deadweight loss from the tax system, while allowing cash transfers to address equity concerns.

Several other consumption-based income tax schemes that work only through the personal tax system have also been suggested. Unlike the two-tiered cash-flow tax, some focus on the uses rather than the sources of income, deducting “nonconsumption” outlays from taxable income but adding back in any drawdown from past saving used to finance consumption. The practical difficulty with these approaches is the definition of “nonconsumption” spending. This could include, for instance, tax preferences for spending on health, education, and state and local taxes, narrowing the tax base. It also could deny deductions for purchasing consumer durables, subjecting, for instance, the purchase of an owner-occupied house to immediate taxation. This is characterized as the “prepayment” of taxes on the future consumption of services provided by the consumer durable. The personal cash-flow tax described in U.S. Department of the Treasury (1977) would give wide latitude to taxpayers on whether they prepay taxes on the consumption of durable goods or pay the tax over time, as long as the present value of the lifetime tax is unaffected.

A comprehensive personal income tax. There is substantial support in the tax-policy literature that a consumption-based tax promotes economic efficiency and growth (see Atkinson and Stiglitz 1976; Chamley 1986; and Judd 1985). The essential factor in this literature is that the focus of the tax on consumption does not distort the relative price of future consumption. Bradford (1986) also argues that the consumption tax does not penalize those who save more. However, pure consumption taxes are not the norm in either developed or developing countries. Banks and Diamond (2010) argue that, in fact, optimal tax theory would allow for the taxation of capital income:

Support by economists and tax lawyers for exempting capital income from direct taxation has been influenced by the well-known Atkinson–Stiglitz and Chamley–Judd analyses. However, we conclude that the policy relevance of the sharp finding of the optimality of no taxation of capital income is thoroughly undercut by the implications of large uncertainty about future earnings and the growing disparity in earnings as a cohort ages. Adding such uncertainty and disparity to the frameworks employed by Atkinson–Stiglitz or Chamley–Judd results in the conclusion that taxation of capital income or of wealth is indeed part of optimal taxation.

Banks and Diamond thus argue for a broader base for income taxes, nuanced by adjustments for the length of time an asset has been held and for the taxpayer’s age. From an equity point of view, since capital income is more unequally distributed than labor income, the taxation of capital can also increase the progressivity of income taxes. However, as Auerbach (2008) has noted, “[T]he literature lacks results suggesting that labor income and capital income should be
treated equally by the tax system, and there is no obvious intuition suggesting that such a result might make sense. It is no accident that the literature has focused on whether it is efficient for capital income taxes to be zero, rather than on whether there should be equal taxes on capital income and labor income.”

Taxing the return on saving raises a number of practical issues, because it is impractical to tax capital income as it accrues, and taxpayers have latitude in the timing of their realization. In particular, taxpayers can realize capital gains whenever they choose. Moreover, an optimal tax system would tax real capital income — that is, capital income adjusted for inflation. Adjustments for inflation are made, for example, in the specification of exemptions and tax brackets, but not typically for the income itself. Finally, because capital income is more mobile than labor income, it is difficult for one country to tax capital at a significantly higher rate than others.

**Corporate Taxes**

Corporate tax systems vary widely across countries and are subject to myriad criticisms. The corporate tax has two distinct roles. First, it can be part of a two-tier cash-flow tax scheme designed to tax consumption. Second, it can be part of a system designed to tax the rate of return on saving, as under a comprehensive personal income tax. Both roles require integration with the personal tax system, but in different ways. In the latter case, the corporate tax could be an alternative to taxing the rate of return at the personal tax level, allowing this return to be taxed as it is earned rather than when it is realized by stockholders, either through dividends or capital gains. In this case, the corporate tax will face some of the same challenges that arise when taxing capital returns in the personal tax system, particularly with respect to adjustments for inflation and problems created by the mobility of corporate income. To avoid taxing corporate income twice, the taxation of dividends and capital gains at the personal level would have to be adjusted for taxes already paid at the corporate level.

Corporate taxes are often viewed as highly progressive, under the assumption that corporations are owned by the wealthy. However, this is not the case. The long-run incidence will depend, inter alia, on the openness of the economy, the structure of production, and the structure and evolution of the tax (Auerbach, 2006). In a small, open economy, the long-run incidence will fall primarily on wages, but wages throughout the economy, not just in the corporate sector (Harberger, 2006). Estimates of the incidence vary widely, but provide substantial evidence that at least one-half of the corporate tax falls on labor, with estimates ranging as high as 100 percent (see, among others, Gentry (2007), Gravelle (2010) and Harris (2009) for recent discussions of the incidence of the corporate income tax).

**Social Insurance Contributions**

Social insurance for pensions, health care, and unemployment is often financed from payroll taxes/contributions. The rationale for this type of financing depends on the type of benefits. For those that are a function of earnings, such as pensions or unemployment compensation, a payroll levy can mimic an insurance premium; it is thus reasonable to consider the levy as a “contribution.” The tax component of a pension contribution is equal to the difference between the actual contribution and what the participant would voluntarily pay, rather than the full
amount of the levy. In this case, a payroll levy can be an efficient mechanism for financing an insurance benefit.

For other social insurance benefits, such as health care, there is no connection between the amount of the levy and access to benefits, although it might be necessary to pay a minimum amount to be eligible for benefits. In this case, almost the entire payroll levy is a true tax, and any increase in the tax resulting from higher wages is a pure tax. For these reasons, the use of a payroll levy for benefits that are unrelated to wages carries all the problems of any wage tax. Distinguishing a special tax for, say, health care can result in higher payroll taxes and more deadweight loss. Moreover, the implications for informality are also of great concern. Finally, earmarking funds where this yields no efficiency gains curtails the need to carefully set priorities and for a program to compete for resources with other programs.

4.3. WEALTH TAXES

Although wealth taxes do not fit into the same conceptual framework as consumption and income taxes, they do fit into the revenue systems in most countries. Several variants are currently used.

First, some countries levy a tax on the stock of wealth. This type of tax shares many of the conceptual and practical problems of taxing the return on saving. The practical problems are especially intimidating: a wealth tax would require a comprehensive measure of wealth, in all its forms, to avoid creating serious distortions. It would repeatedly tax savings, as would a tax on the return to saving. As noted by Mirrlees et al. (2012, p. 347), “where attempts have been made to levy a tax on a measure of current wealth — in France, Greece, Norway, and Switzerland, for example — practical experience has not been encouraging.”

Second, many countries tax the transfer of wealth. The argument for such a tax is that it promotes equality of opportunity — that people who start off with inherited wealth have an inherent advantage. This rationale suggests the tax should be levied on the recipient rather than the donor, which is not typically the case. When considering a tax on wealth transfers, it is important to determine whether charitable contributions are exempt from income taxes. In the United States, for instance, the top marginal rate for the inheritance tax is 40 percent. In that case, a charitable contribution from a wealthy donor is worth two-thirds more to the recipient as a gift or bequest to a family member, with the public making up the difference in the form of lost revenue. One can question the degree to which the charitable gift exemption has resulted in, say, universities stockpiling large endowments. Absent the charitable exclusion, tax revenue would likely increase. The question then would be whether the nonprofit recipient of charity or the government would make better use of the funds.

Third, many jurisdictions levy a tax on real property, which can take the form of land; improvements to land (buildings and other immovable objects); or personal property (movable items of property, such as automobiles or furniture). Since land is (roughly) fixed in quantity and immobile, it is an efficient tax base. Moreover, the taxes can be viewed as payment for the infrastructure services available to the land. In this sense, a tax on land can (roughly) satisfy the “beneficiary pays” criterion for tax fairness. However, if the tax is too high, it can become confiscatory. It will discourage the development of new land and have an unfair impact on those
who hold wealth in real property. If property taxes are imposed at the local level, high tax rates will also drive capital into lower tax jurisdictions — that is, into jurisdictions where the tax is lower relative to the infrastructure and amenities provided by the local government. Taxes on improvements to land and personal property are more distortionary and less easy to justify. However, one could construct a tax on houses that would correlate with a tax on housing services in cases where the consumption of owner-occupied housing would otherwise escape taxation, as with most VAT systems.

### 4.4. ADDITIONAL REVENUE POLICY ISSUES

**Efficiency.** All things equal (never the case in practice), the deadweight loss created by a 38 percent tax rate (the average revenue to GDP ratio in high-income economies) will be more than three times the deadweight loss created by a 21 percent tax rate (the average revenue to GDP ratio in low-income countries). Of course, all things do not have to be — and typically will not be — equal. Some taxes are much more efficient than others because they have less effect on behavior. Some carry no deadweight loss at all. The Organisation for Economic Co-operation and Development (Johansson et al. 2008) identified which taxes are most conducive to growth. In order of their impact on growth, the study considered (1) property taxes, (2) consumption taxes, (3) income taxes, and (4) corporate taxes.

**Equity.** The equity of a tax system is evaluated using two criteria. First, people who are equally situated should be treated equally — horizontal equity. Two households with the same income and characteristics should pay the same taxes. Second, the level of taxes paid should vary directly with the ability to pay — vertical equity. What constitutes ability to pay, however, is a matter of opinion. Adam Smith argued for proportional taxes; many would argue for progressive taxes. However, to measure progressivity, it is necessary to specify (1) a measure or taxes — for instance, gross taxes or taxes net of benefits, (2) a measure of welfare against which to compare taxes paid — for instance, income or consumption, (3) a time period — for instance, one year or a lifetime, and (4) a definition of a taxpayer — for instance, an individual or a family. The last category is perhaps the most problematic because family structure varies over a lifetime. The measured progressivity of all of the taxes discussed above will vary according to the definition of progressivity, but all admit to designs that can affect progressivity. However, designing progressivity into the tax system in isolation is less efficient than using both the tax and expenditure systems to achieve desired progressivity in net taxes.

**Earmarking tax revenue.** To insulate a program from continuous reevaluation of its priority and to ensure a smooth revenue stream, advocates of the program will often argue that it should have a dedicated (earmarked) revenue source. This approach is valid in two cases: when the revenue stream comes from user charges (say, road or bridge tolls) or the dedicated revenue stream is in the nature of an insurance premium (say, payroll levies to finance pension benefits or unemployment insurance, as discussed above). In these cases, earmarking internalizes the cost of the benefits directly to the beneficiary, bringing quasi-market forces to bear on government programs. To be efficient, these earmarked charges/premium contributions must be carefully structured so there is a close tie between the charges and the benefits they finance. Otherwise, they can be perceived as taxes, with the concomitant inefficiency they engender. Although such links can exist for pensions or unemployment insurance, they do not exist for health care, even
when the financing source is characterized as a social insurance contribution (see above). As a result, some countries are considering shifting the financing of social benefits — such as health care — that do not vary with income away from payroll taxes and toward consumption taxes. Absent imperfections in tax administration, the two taxes would have roughly the same incidence in the long run. In practice, however, it may be easier to capture the consumption tax base. (See IMF (2012b) and Gillingham and Jousten (2014)).

**Intergovernmental fiscal relations.** In many countries, responsibilities for government functions are allocated to different levels of government. A general principle in this allocation is that responsibility should be given to the level of government closest to the citizens being served that can still benefit from economies of scale in implementing the program. Similarly, responsibilities for collecting revenue are also shared among central and subnational levels of government. Thus, mechanisms must be developed to enable the resources to flow to where they are used. This will apply to health care, where, for example, some facilities should be local and others more centralized. The resources must flow to the appropriate level of government, not only to allow the programs to function, but also to ensure that poor regions of a country are not too disadvantaged in obtaining needed resources.

**Tax revenue constraints.** As a government tries to raise ever more revenue to implement more and more policies, the deadweight loss/excess burden of raising that revenue increases at an increasing rate. This fact reduces the probability that another government program will (1) have positive net benefits, and (2) be the best use of scarce resources. This is why it is so important to set spending priorities. As the tax burden increases, it becomes more difficult for a potential use of government resources to have positive MNSB.

### 4.5. NONTAX REVENUE

Not all revenue comes from taxes. Nontax revenue can come from a variety of sources, with significantly different impacts on efficiency. Nontax revenues include such items as the following:

- Grants from foreign governments or other donor organizations;
- Profits of state-owned enterprises, including profits from state oil companies (losses reimbursed from the budget would be recorded as expenditures on transfers);
- Fees for the issuance of permits and licenses, such as vehicle registration and building permits;
- Rents, concessions, and royalties from contracts that allow private entities to use government property, including the extraction for profit of natural resources;
- User charges for goods and services provided by the government, such as bridge tolls, admission to museums, and tuition for government-run schools;

**Official development assistance (ODA).** Nontax revenues are typically small in comparison to tax revenues, with two exceptions. Some LICs rely heavily on grants to finance their governments. Benedek et al. (2012) constructed a new database on tax revenues and ODA that
shows that ODA averaged 4.7 percent of GDP from 1980 to 2009, compared to 16.1 percent of GDP for tax revenue in a sample of 118 low- and middle-income countries. The share varied widely, however, ranging up to 96 percent of GDP. The share was highest during the late 1980s and early 1990s; more recently, it averaged roughly 4 percent. The level of ODA — in particular grants, which have accounted for more than 90 percent of ODA — is negatively correlated with domestic tax collections. This result is not surprising, since the added revenue allows governments to increase their overall spending, presumably reducing its marginal benefit.

Proceeds from the sale of natural resources. Nontax revenues are also important in countries rich in natural resources. However, there are serious conceptual issues with the classification of the proceeds from natural resource exploitation, both for measuring GDP and defining current revenue in the budget. According to widely accepted national income accounting standards, the full proceeds from extracting natural resources are included in GDP. Arguably, however, the value of the asset before extraction — for example, the amount that a willing buyer would pay for oil in the ground — is a government asset. If we accept this premise, only the value added of extracting the asset — for instance, drilling costs — should be counted as a current output flow. This treatment would have a significant impact on measured GDP for resource-rich countries. For instance, Qatar had the highest GDP per capita (Purchasing power parity basis) in the world in 2012. Much of this GDP, however, reflected the drawdown of its stocks of natural resources.

Similarly, the proceeds for government from natural resource exploitation — either from state-owned enterprises or contracts with private entities — are counted as current revenue. Arguably, only the revenue generated by extraction should be counted as revenue, and the remainder counted as a sale of assets providing a source of financing. This treatment would mirror that used in the government budget when a public asset is sold or “privatized.” The use of these resources is equivalent to a tax on property, and, depending on how society allocates property rights to the natural resources — do they belong to anyone? to the current generation? to current and future generations? — the incidence of the tax can vary widely.

Equity issues for nontax revenues vary by source. For ODA, there is no tax incidence; the distributional impact depends on how the resources are used. As discussed above, on the other hand, natural resource proceeds are more akin to a tax. Because the property rights to the natural resources are problematic, the distributional incidence of the implicit tax is extremely complicated. Resource revenue is not a free good, however. When it is used to finance current spending, the burden of the tax falls on future as well as current generations. The equity of raising resources in this fashion will depend on how society weights the welfare of members of current and future generations.

4.6. VARIATIONS IN REVENUE SOURCES BY DEVELOPMENT LEVEL

Figure 1.7 displayed how revenue levels, as a share of GDP, have varied across countries by level of development. Table 1.2 summarizes how the average share of GDP from each revenue source varied, on average, from 1980 to 2010. Because total taxes are much higher as a share of GDP in high-income OECD countries, the components also represent a higher share of GDP. High-income non-OECD countries have a significantly different distribution of revenue sources because of the importance of resource rents.
V. FISCAL SUSTAINABILITY

Budget Balance

Expenditure and revenue need not be equal on a continuing basis. In fact, *countercyclical fiscal policy* allows for an increase in outlays and a reduction in revenue during downturns and the opposite during upturns. Over long periods, however, expenditures less revenue — the fiscal surplus — should be at least zero, or else the government will accrue significant debt. In public discourse, attention is typically paid to the overall budget surplus (figure 1.8). From 2000 to 2010, governments typically ran overall deficits (negative surpluses). The share of budgets with overall deficits ranged from 62 percent for high-income countries to 86 percent for low-income countries (71 percent for all countries combined). The median overall deficit in 2010 ranged from 3 percent in low- and lower-middle-income countries to 5 percent in high-income countries. These deficits were abnormally high, however, especially in high-income countries, in the wake of the global recession. In 2000 and 2010, for example, the median overall deficits across all countries were 1.8 and 1.6 percent, respectively, compared to 3.6 percent in 2010.

However, from the standpoint of fiscal sustainability, the *primary* budget surplus — equal to the overall surplus plus interest outlays — matters more. If a country can cover its new spending with current revenue, and GDP grows at close to the rate of interest it pays on its existing debt, the debt-to-GDP ratio will change little. With a modest primary surplus, the debt-to-GDP ratio can fall even if the budget is in overall deficit.

---

**TABLE 1.2 DISTRIBUTION OF REVENUE SOURCES BY LEVEL OF DEVELOPMENT, 1980–2010**

<table>
<thead>
<tr>
<th>Revenue sources as a percent of GDP</th>
<th>Low income</th>
<th>Lower-middle income</th>
<th>Upper-middle income</th>
<th>High income (non-OECD)</th>
<th>High income (OECD)</th>
<th>All countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government revenue, total</td>
<td>18.4</td>
<td>26.4</td>
<td>28.5</td>
<td>33.8</td>
<td>41.5</td>
<td>28.7</td>
</tr>
<tr>
<td>Government revenue exc. grants</td>
<td>15.2</td>
<td>25.6</td>
<td>27.9</td>
<td>33.7</td>
<td>41.4</td>
<td>27.8</td>
</tr>
<tr>
<td>Government taxes, total</td>
<td>13.0</td>
<td>17.7</td>
<td>20.7</td>
<td>15.7</td>
<td>35.4</td>
<td>20.5</td>
</tr>
<tr>
<td>Income taxes, total</td>
<td>3.5</td>
<td>5.0</td>
<td>5.4</td>
<td>5.9</td>
<td>12.9</td>
<td>6.9</td>
</tr>
<tr>
<td>Corporate income taxes</td>
<td>2.2</td>
<td>2.9</td>
<td>3.4</td>
<td>2.4</td>
<td>3.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Individual income taxes</td>
<td>1.5</td>
<td>1.9</td>
<td>2.3</td>
<td>2.8</td>
<td>9.7</td>
<td>4.7</td>
</tr>
<tr>
<td>Taxes on goods and services</td>
<td>5.0</td>
<td>6.1</td>
<td>7.1</td>
<td>5.1</td>
<td>11.2</td>
<td>7.4</td>
</tr>
<tr>
<td>VAT revenue</td>
<td>4.9</td>
<td>5.0</td>
<td>5.2</td>
<td>6.2</td>
<td>6.8</td>
<td>5.9</td>
</tr>
<tr>
<td>Trade tax revenue</td>
<td>3.7</td>
<td>4.9</td>
<td>4.6</td>
<td>2.7</td>
<td>0.6</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Memorandum items:

<table>
<thead>
<tr>
<th></th>
<th>Low income</th>
<th>Lower-middle income</th>
<th>Upper-middle income</th>
<th>High income (non-OECD)</th>
<th>High income (OECD)</th>
<th>All countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of countries</td>
<td>37</td>
<td>48</td>
<td>41</td>
<td>18</td>
<td>30</td>
<td>174</td>
</tr>
<tr>
<td>Corporate income tax rate</td>
<td>39.0</td>
<td>33.5</td>
<td>33.3</td>
<td>33.8</td>
<td>33.8</td>
<td>33.5</td>
</tr>
<tr>
<td>VAT rate, 2010</td>
<td>16.2</td>
<td>13.8</td>
<td>15.7</td>
<td>18.0</td>
<td>18.0</td>
<td>15.8</td>
</tr>
</tbody>
</table>

Source: International Monetary Fund (2011)

Note: Availability of data varies by source, country and year, so percentages do not add to higher-level categories.
Countries have been more successful at running primary surpluses over the past decade than at running overall surpluses (figure 1.9, although data on primary deficits are available for far fewer countries). Fifty-one percent of the budgets for which primary deficits are available from 2000 to 2010 were in primary surplus (57 percent before the recession began). This share ranged from 44 percent in upper-middle-income countries to 58 percent in low-income countries. The IMF’s *World Economic Outlook* projects that the level of primary surpluses/deficits as a share of GDP will rebound from the global recession by the middle of this decade.
Accumulation of Debt

As noted above, a country need not run an overall budget surplus to reduce its debt-to-GDP ratio. Moreover, even if a country runs a series of primary deficits, its debt-to-GDP ratio need not explode. However, with realistic assumptions about interest rates and GDP growth, a country will have to run primary surpluses to reduce its debt-to-GDP ratio. In fact, many countries have reduced their gross debt as a percentage of GDP over the past decade even with the fiscal pressure created by the recession (figure 1.10, although the performance in low-income countries is shaped, in part, by debt relief). The median gross debt-to-GDP ratio in 2010 ranged from 37 percent of GDP in low-income countries to 48 percent in high-income countries. Again, however, the variation within groups is very large. A number of countries, including vulnerable ones such as Greece and Italy, have debt-to-GDP ratios over 100 percent, and the IMF has estimated that the ratio in the largest economy — the United States — reached 100 percent in 2011. Arguably a better indicator of a country’s indebtedness is the ratio of its net debt to its GDP. This is especially true for countries, such as Singapore, that have significant gross debt that is more than offset by their financial assets. Unfortunately, however, net debt estimates are not available for many countries. In addition, determining what should count as an asset — for instance, government lands in Alaska — is subject to debate, and valuing many assets can be difficult.

Assessing Fiscal Sustainability

To determine whether fiscal policy is sustainable, we start by defining the government’s borrowing requirement (see Croce and Juan-Ramón [2003] for a detailed discussion):

\[ BR_t = D_t - D_{t-1} = i_t D_{t-1} - PS_t, \] (1.3)
where \( BR \) is the borrowing requirement, \( D \) is the level of debt, \( i \) is the rate of interest the government pays on debt, \( PS \) is the primary surplus, and the subscript identifies the time period. In words, the borrowing requirement is equal to the change in the level of debt minus the primary surplus. Dividing equation (1.3) by \( GDP_t \), it is straightforward — although perhaps tedious — to derive the debt dynamics equation:

\[
d_t = \beta_t d_{t-1} - ps_t, \tag{1.4}
\]

where \( d \) is the debt-to-GDP ratio; \( \beta \) is the discount factor, defined as \((1+r_i)/(1+g_i)\); \( r \) is the real rate of interest; \( g \) is the real rate of growth in GDP; and \( ps \) is the ratio of the primary surplus to GDP; and the subscripts identify the time period (note that \( d_{t-1} \) is the ratio of the debt in period \( t-1 \) to GDP in period \( t-1 \)). Again in words, the debt-to-GDP ratio in period \( t \) is equal to the debt-to-GDP ratio in the previous period times a discount factor, minus the ratio of the primary surplus to GDP.

The key variable in the debt sustainability analysis is \( \beta \), which is equal to the factor by which debt grows (1 plus the interest rate), divided by the factor by which GDP grows (1 plus the growth rate). If the interest is high relative to the growth rate, debt will grow rapidly before any primary surplus offset. Conversely, if the rate of GDP growth is high relative to the rate of interest, debt will fall. Unfortunately, it would be a mistake to assume that growth will be faster than the rate of interest for a long period. In fact, a steady state in which \( g \) is greater than \( r \) is dynamically inefficient, meaning that current generations can save less and consume more — reducing GDP growth — without reducing the welfare of future generations. (See Barro [1976] for a discussion of dynamic inefficiency. See Phelps [1966] for a classic exposition of related growth concepts.)

From equation (1.4) we can solve for the primary surplus needed to achieve any level of debt, conditional on the growth rates. For instance, to keep the debt-to-GDP ratio constant from one period to the next, the primary surplus must be the following:

\[
ps^* = dt \left[ \frac{r-g}{1+g} \right]. \tag{1.5}
\]

As equation (1.5) demonstrates, the primary surplus has to equal the lagged debt ratio times a factor equal to the rate of interest less the rate of growth of GDP “discounted” by 1 plus the rate of growth in GDP. If the interest and growth rates are equal, the debt ratio will remain constant with a zero primary surplus. In general, debt sustainability is easier (1) the larger the rate of growth, and (2) the lower the rate of interest.

Equations (1.3) through (1.5) can be manipulated in many ways to determine the primary surplus path that will achieve any debt target over any horizon. The exercise can also allow for systematic and random changes in \( r \) and \( g \). The point of this discussion, however, is simply to lay out the intuition behind evaluating fiscal sustainability and developing policies to achieve surplus and debt ratio targets. These exercises reinforce the importance of well-designed expenditure policies and a fair and efficient tax policy to allow the government to achieve its goals without creating fiscal stress. Although tax policy is largely separable from expenditure policy, it is true that government spending is constrained by the cost of raising resources. It is not enough to have a “good” policy. Rather, that policy must entail the best use of scarce resources.
VI. IMPLICATIONS FOR HEALTH POLICY

The fiscal policy framework discussed above can be used to formulate health care policies that (1) address society’s equity goals, (2) use resources efficiently, and (3) are fiscally sustainable. If we follow the sequencing laid out above, the implications of a well-structured fiscal policy for health care are the following:

- **Decide on the appropriate role of government in the health sector.** Several options exist, ranging from direct provision (for instance, the United Kingdom) to single-payer financing (for instance, Canada) to regulating managed competition (for instance, Switzerland or the Netherlands). The push for universal health insurance has tended to increase the role of government. However, universal health insurance need not imply universal access to health care. Many countries with universal health insurance must adopt nonprice rationing, (including by queue) to control costs, with implications for access. In deciding the role of government, the choices made must be carefully and systematically justified. For example, is the government addressing the lack of access for the poor and vulnerable or the implications of market imperfections for the general population? How can the government fulfill its role most efficiently? Consideration should be given to adopting market-based mechanisms — price rationing — where possible.

- **Recognize that health spending must compete with other uses of government resources.** It is natural for a policy maker in a particular sector to believe that sector is underfunded. However, advocates for other sectors will reach the same conclusion for their sectors. A well-designed budget process will require each health care program to demonstrate that the resources requested will be more productive in health than in other sectors (allocative efficiency).

- **Maximize both allocative and productive efficiency within the health sector.** This involves not only decisions about the role of government (see first bullet), but also (1) efficient allocation of resources across public health care programs (for instance, among primary, secondary, and tertiary care), and (2) efficient production of the health care products and services for which the government takes responsibility. Before seeking additional resources for government health care programs, policy makers should ensure the resources already dedicated to the sector are used efficiently.

- **Recognize long-term challenges.** Technological change in the health sector is rapid, with implications for which health care services can be provided and which services are likely to be demanded by patients. Moreover, population aging will add to demand pressures. Health care policy makers must account for these dynamics in planning both the current role of the government in the health sector as well as planning for how that role will evolve.

- **Raising the resources for government health care programs is largely outside the purview of health care policy makers.** To the extent that health care services are provided as government transfers, raising revenue should be left to the overall budget process, with tax policy experts designing the most efficient procedures for raising revenue (or borrowing). All government programs should compete for resources, and priorities should be continually reevaluated and adjusted in response to changing circumstances. Earmarked ODA can counter this goal if (1) it is greater than the amount the country would otherwise spend on
government health care, or (2) it imposes an allocation of resources among programs that would otherwise be inefficient.

- **User charges could be considered where they would improve efficiency.** To the extent that government health care programs can use market forces and prices to ration health care, these proceeds could indeed be recycled directly into the health care sector. There is understandable concern that user charges in the health sector not subject citizens to undue financial risk. However, user charges that are systematic and subject to very low variance — for example, the cost of routine annual physicals — would not present such a risk and permit more effective rationing of health care (of course, the government could subsidize user charges for households that cannot afford them).

- **Fiscal sustainability is paramount.** The health care policy goal must be to provide the best health care possible, but only within the resource constraints consistent with fiscal sustainability.
REFERENCES


Ebrill, Liam, Michael Keen, Jean-Paul Bodin, and Victoria Summers, ed. 2001. The Modern VAT. International Monetary Fund: Washington, DC.


———. 2012a. World Economic Outlook, October database.


ANNEX 1. ALLOCATIVE AND PRODUCTIVE EFFICIENCY

1.1. ALLOCATIVE EFFICIENCY

Resources Should Go to Their Highest Use

Allocative efficiency is achieved when (scarce) resources are put to their highest use. In other words, an economy — or a sector of an economy — is allocatively efficient if there is no shift of resources among sectors or subsectors that can increase social welfare. To achieve allocative efficiency, two key conditions must be met (note that we are talking here about factors of production; the concepts of money and financing are separate):

- First, resources must be allocated efficiently across production/consumption sectors as a whole. The total resources allocated to a sector must be optimal.
- Second, within a production/consumption sector, resources must be allocated efficiently between the public and private entities operating in that sector.

The ordering of these conditions is not important, since they must be met simultaneously. It would be equivalent to say that total resources must be allocated appropriately between the public and private sectors, and that, within each of these sectors, resources must be allocated efficiently among production/consumption sectors. Allocation of resources between the private and public sectors brings us back to the fundamental question raised at the outset: What should the role of the government be in the economy? Again, as discussed above, this role will vary across production/consumption sectors. The government will have a much larger role to play in national defense than it will in agriculture.

The basic idea underlying allocative efficiency is that resources should go to their highest use. The last dollar of spending should have the same marginal net social benefit (MNSB), where MNSB is defined as the social benefit produced by the spending less the cost of the resources used. Conceptually, this should hold across all activities, in both the public and private sectors. Efficient allocation of resources will determine the overall size of the government — the level of expenditure — as well as the allocation of resources across the functions and programs of the government.

Determining “Highest Use”

Within a diversified, private-sector firm, allocative efficiency is a positive concept. The firm has a straightforward objective function — maximize profits — with an obvious metric. However, private markets identify only private benefits and private costs. The government, acting as an agent for society as a whole, is concerned with the level of social welfare — that is, social benefits and social costs. Since social welfare is inherently normative, there is no unique metric for evaluating alternative allocations of resources. Even in a democracy, it is not easy for the preferences of individuals to be effectively represented in an election. As a result, allocating resources to reflect social preferences is as much an art as a science.
**A Practical Approach**

In the absence of an unambiguous behavioral rule for policy makers, the emphasis has to be on common sense. First and most important, decisions must be made on the appropriate scope of governments within production/consumption sectors as producers, financers, and regulators. These decisions should reflect and address the issues discussed in the previous section on the role of government. Does the sector produce a public good? How imperfect are markets in the sector? Will private markets yield an inequitable outcome? To the extent that unfettered private markets yield outcomes that are socially desirable, the government need not play an intrusive role. To the extent that market forces can be brought to bear even on government activities, it is easier to assess allocative efficiency.

**1.2. PRODUCTIVE EFFICIENCY**

Productive efficiency can be defined in several ways: (1) minimize the cost of producing a given level of output; (2) maximize the output produced with a given level of inputs; or, most generally, (3) maximize the difference between the value of output and the cost of inputs. It is a necessary condition for Pareto efficiency (under which no member of society can be made better off without making someone else worse off), which also encompasses efficiency in consumption. Evaluating productive efficiency is straightforward for most competitive firms because outputs and inputs are well defined, and money serves as a convenient metric.

For government programs, outputs are typically not well defined, nor is a money metric available. For a long time, the output of a government program was inferred from the level of inputs — in education, for instance, from the number of teachers, classrooms, and books. More recently, the focus has shifted to outputs — for instance, the number of students receiving an education. Better still would be a focus on outcomes — for instance, the level of education achieved by students. When the outcomes entail social as well as private benefits, assessing outputs and outcomes is difficult; again, decisions must perforce be based on incomplete information and a particular set of social preferences.

**1.3. TYPES OF GOVERNMENT INTERVENTION**

Government intervention in the economy can take several forms. From least to most intrusive, the government can do the following:

- regulate private-market activities, either directly or implicitly through the imposition of taxes and subsidies;
- finance the production of goods and services by contracting with private-sector entities; or
- directly produce and distribute goods and services.

As economies develop and create effective governance institutions — for instance, clear property rights and well-functioning contract laws — the need for governments to directly produce goods and services diminishes, and it becomes more feasible to adopt less intrusive means to achieve
government objectives. This, in turn, will also improve the ability of policy makers to increase productive — and allocative — efficiency. Politicians constantly promise to eliminate “waste, fraud, and abuse” and thereby reduce the cost of government. First it will be impossible to eliminate all waste, fraud, and abuse, especially if inefficiencies stem from political decisions. Second, a reduction in waste, fraud and abuse is a once and for all improvement in the level of costs. More important is to address the dynamics of government spending and ensure that inefficient trends do not creep into fiscal or health policy.
ANNEX 2. THE ROLE OF THE STATE IN THE ECONOMY

Why should the government intervene in the market economy? Although there is no definitive answer to this question, a number of goals are generally accepted. The variation in the approaches followed by different countries revolves primarily around how best to achieve these goals.

Overarching objective. From a utilitarian perspective, the goal of government intervention is to improve the welfare of society. On a somewhat more practical level, this objective translates into promoting an appropriate mix of two derivative goals: efficiency and equity. Improvements in efficiency increase the level of output. Improvements in equity, on the other hand, allow that output to be distributed in a manner that society views as more fair. The first of these objectives is positive — it entails no value judgments — while the latter is inherently normative. Weighing the two objectives implies either an implicit or explicit social welfare function. The problems with defining social welfare functions are well known and beyond the scope of this note. In what follows, we simply assume that society has a mechanism for choosing among alternative policies.

2.1. PARETO EFFICIENCY

The first fundamental theorem of welfare economics states that, under certain conditions, any competitive equilibrium will lead to a Pareto efficient allocation of resources (no one can be made better off without someone else being made worse off). To simplify, a market outcome will be Pareto efficient if the market is characterized by a very large number of buyers and sellers who have perfect rationality and perfect information. For better or worse, these conditions do not always apply, and there are several circumstances in which the need for government intervention to enhance efficiency is widely accepted:

Public Goods

A pure public good is one that is both nonexcludable in that an individual cannot effectively be excluded from consuming it and nonrivalrous in that consumption by one individual does not reduce the availability of the good to other potential consumers. Because no party can be excluded from consuming the public good, no private entity can recapture the cost of providing it. Consequently, government intervention is required.

Classic examples of public goods include national defense, lighthouses, and air. The last of these occurs in nature and requires no government intervention to ensure its availability (of course, the discussion would be different if we tried to classify “clean air” as a public good). Public goods that do not occur in nature — national defense and lighthouses in the examples above — require government action to ensure their supply. The government may choose to provide the service itself, as is often the case for national defense. Some might argue that national defense is not a public good — for instance, if a government builds special purpose defensive installations to protect only some of its citizens. This point is well-taken, but this defense strategy arguably just creates a public good for a subset of a country’s citizens. Even a defense installation specifically for, say, New York City will benefit all of the citizens in a manner that is both nonexcludable and nonrivalrous. Public goods can take many forms, including within the health sector. For
instance, some public health projects, such as draining swamps, provide benefits that are nonrivalrous.

It need not be difficult to assign the benefits that a public good produces and to impose the costs on those beneficiaries. For instance, before the advent of more sophisticated navigational tools, the benefits of many lighthouses accrued to the ships entering a harbor. The costs of providing the services could, consequently, be recouped as part of harbor fees. As discussed below, a government has a variety of options to ensure the supply of public goods. It does not, however, have the option of allowing a market to work, since no party is willing to provide a good for which it cannot charge.

The key factor that requires government intervention is nonexcludability. Even if a good is nonrivalrous, a market for it can evolve as long as the producers of the good can exclude those who do not pay from consuming the good. Examples include satellite television, whereby the providers of the service can limit access to those who pay despite the service being nonrivalrous; and patents, which allow the producers of new innovations to at least limit the degree to which consumers have unfettered access. A slightly different problem occurs if a good is nonexcludable and rivalrous, such as the clean air mentioned above. Government must intervene to ensure that overuse or misuse of the resource does not unduly curtail its supply. In this case, the intervention is more likely to be regulatory, to avoid what ecologist Garrett Hardin entitled the “tragedy of the commons.”

**Market Imperfections**

Governments often intervene in markets to improve the efficiency with which they operate. The causes of market inefficiency vary. The most important examples are natural monopoly, externalities, and information asymmetry.

**Natural Monopoly.** In most markets, firms operate under a technology in which the long-run average cost of production is achieved with firms — or at least individual production facilities — that are small relative to the size of the market for the good or service being produced. Over some range, larger production facilities operate more efficiently than smaller ones. It is difficult to produce an automobile in a plant with few workers and limited capital. However, as the scale of a plant increases, the economies of scale are exhausted. At some point, it becomes more efficient to build another plant than continue to expand in the same location. This type of technology is conducive to competition, since new firms can enter a market as long as they can operate on an efficient scale. Although some government regulation must exist to ensure that firms compete fairly, market regulation is minimal.

Some markets, however, are characterized by increasing returns to scale over such a broad range that it is more efficient for production to be concentrated in a single firm — *a natural monopoly* — that can achieve the lowest long-run average cost of production. The most common examples of (potentially) natural monopolies are utilities, such as electricity, telecommunications, and water, in which a vast infrastructure is necessary to distribute the good or service. Absent special government intervention, a natural monopolist can use its market power to extract “economic rents,” that is, returns to its investment that are above returns available in “normal” markets.
Governments have addressed natural monopolies in a variety of ways. In the extreme, they can take over the markets themselves, for instance, by directly producing the good or service within the public sector. In this case, the cash flows of the public sector firm are reflected — albeit not always explicitly — in the government budget. Alternatively, governments can regulate the prices that natural monopolists can charge for their output. More recently, governments have promoted competition even in markets in which long-run average costs fall with the size of the firm. For instance, it could be argued that mobile telephone services are a natural monopoly, since spreading substantial fixed or quasi-fixed costs over an ever larger set of customers reduces long-run average cost. However, mobile phone companies are encouraged to compete in the same market in the belief that the gains from such competition outweigh the limited benefits of larger scale.

**Externalities.** An *externality* occurs when there are costs or benefits that (1) are not captured in the price of a good or service being traded, and (2) accrue to parties who are not involved in a transaction. Absent intervention, the price will reflect only the *private* costs and benefits. In addition, a transaction can create *social* costs or benefits. Perhaps the most obvious externalities — in this case social costs — relate to pollution. The buyers and sellers of fossil fuels concern themselves only with the private costs of production and consumption. One individual, consuming gasoline, has no impact on pollution levels, so there is no social benefit to curtailing individual consumption. Consumers taken together, however, do affect pollution levels; so there is a strong argument for government action that will impose these aggregate costs to ensure efficient pricing. Within the health sector, social externalities would argue for the subsidization of immunizations or the provision of clean water and sewer facilities.

The goal in regulating production and consumption externalities is to equalize social costs and benefits, rather than just private costs and benefits. For instance, if a power plant produces pollution — a social cost — at the same time that it produces electricity — a private benefit — the net gain to society is the private benefit less the social cost. Operating without constraint, the power plant and electricity consumers will settle on a price that does not reflect the social cost. Consequently, the government can place restrictions on the market that internalize the social costs to either the producer or consumer. For instance, the government can impose a tax on electricity — known to economists as a *Pigouvian* tax after the economist who suggested it — exactly equal to the social cost of the pollution. It could also impose quantity restrictions that would have the same effect on the amount of electricity produced. Although the imposition of a Pigouvian tax or restrictions on production are straightforward in a world without transactions costs and no other taxes, their application in practice is complicated, as discussed in section 4.1.

**Information Asymmetry.** Information asymmetry occurs when one party to a transaction has more or better information than the other party. This can cause problems with the way in which the market functions, including moral hazard, adverse selection, and information monopoly. Kenneth Arrow (1963) was one of the first economists to discuss the problems of asymmetric information, focusing on the market for health care. Since that time, many economists have contributed to the literature on this topic, including George Akerlof, Michael Spence, and Joseph

---

4. See note 1.
Stiglitz, who jointly won the 2001 Nobel Prize in Economics for their contributions. Since information asymmetry takes many forms, the policy prescriptions are varied, including both adaptations within private markets and public policies to mitigate the costs of information asymmetry.

2.2. REDISTRIBUTION

The first fundamental theorem of welfare economics focuses only on Pareto efficiency. It is not violated by an outcome in which the distribution of welfare is very unequal, although any redistribution will create winners and losers. However, society need not accept an outcome with a distribution of welfare that it views as unacceptable. Rather, the concept of efficiency can be expanded to focus on social welfare; social welfare need not be maximized even if an economy is Pareto efficient. This expansion of the definition of efficiency, however, begs a number of questions. How do we determine what “society” will accept? How does society rank outcomes? Should the focus be on the absolute level of welfare of the poorest in society or on the relative level? Specific answers to these questions are well beyond the scope of this primer. However, it is important that the answers — explicit or implicit — should drive decisions on the role of government in redistributing welfare. These answers redefine efficiency in terms of maximizing social welfare instead of simply meeting technical conditions for Pareto efficient private production and consumption. Consequently, they allow for winners and losers, providing a basis for deciding whether the social value enjoyed by the winners is greater than the social value suffered by the losers. Pareto efficiency is positive; the maximization of social welfare is inherently normative.

A key question that must be answered in designing redistributive policies is whether there is a trade-off between the level and distribution of welfare. To the extent that the fiscal effort required to redistribute welfare — that is, tax and transfer programs — reduces growth, these programs can be self-defeating. They may improve a household’s relative welfare, but they can, at the same time, reduce its absolute welfare. A balance is necessary between the increase in welfare for poorer households and any possible negative impact on growth and aggregate welfare.

A wide variety of policy instruments are used to redistribute welfare. At one extreme, cash can be redistributed to poor households, leaving the household discretion on how best to use the additional resources. On the other hand, redistribution can be effected through a variety of in-kind transfers. These can include, among others, food distribution programs, housing subsidies or provision of public housing, and programs to provide access to health care services. Moreover, the welfare benefits can be conditional on, for instance, keeping children in school and ensuring receipt of preventive health care.

The government might also supplant private markets in the provision of certain goods and services — in particular, health and education. Justifications for this role include the following:

- Subsidies to poorer households to access private providers could have unintended consequences. Dual markets could develop in which richer households have access to better health and education services, and the quality of services provided by private markets to subsidized households will be substandard.
The social benefits of a healthy and well-educated population are greater than the private benefits accruing to those consuming the services — in other words, broad-based access to quality health and education services have positive externalities — and government provision is the only way to achieve these social benefits.

Note that these justifications are similar to the reasons given for special treatment for what Richard Musgrave (1957, 1959) called *merit goods* — that is, goods that individuals or society as a whole believe should be supplied on the basis of need, rather than ability or willingness to pay. As discussed below, even in a situation where the government is the primary regulator and financer of health and education, it has options about how best to achieve its goals, with concomitant variation in the degree of intrusiveness.
ANNEX 3. EXCESS BURDEN OF TAXATION

In the absence of taxes, markets “clear” — that is, demand is equal to supply — at a price that is Pareto efficient: any increase or decrease in price will reduce total welfare for the producers and consumers of a good (including, also, labor as a good). To provide the resources necessary to finance the government, however, taxes must be imposed. As a result, the welfare from the exchange of the taxed good is reduced, to be offset by the increase in welfare yielded by what the government does with the revenue. The policy lesson is exactly that which has been emphasized above. The uses of the resources by the government must have higher social benefit than the loss in welfare from the imposition of the tax.

The loss in welfare from the imposition of a tax can be easily demonstrated graphically. Figure 1.11 depicts a market without taxes. The downward sloping demand curve (consumer demand decreases as the price increases) crosses the upward sloping supply curve (producers increase their supply as the price increases) at a price of 50 and a quantity of 100. At this equilibrium, consumers enjoy a “surplus” equal to the shaded area between demand curve and the equilibrium price line (constant at 50). This arises because consumers would have been willing to pay a price of 90 for the first unit of the product they consumed, 70 for the 50th unit, but only 50 for the 100th unit. However, they were able to buy all the units they consumed at a price of 50, so the pleasure they got from consuming the first 99 units was much higher than the price they had to pay. Similarly, producers enjoy a surplus equal to the shaded area between the equilibrium price line and the supply curve. The total surplus is equal to the sum of the two shaded areas.

If a tax of 15 is imposed on sales in the market depicted in figure 1.11, the equilibrium in the market shifts. The supply curve shifts up by 15, because the suppliers must receive an additional 15 per unit to receive the same net proceeds. The higher supply curve yields a new market equilibrium with a price of roughly 57.5 and a quantity of slightly over 80. In this new equilibrium, the rectangle shaded in purple is equal to the revenue yielded by the tax (tax amount times quantity). The top half of this area represents a reduction in consumer surplus, and the bottom half represents a reduction in producer surplus. However, the lost consumer and producer surplus...
surplus exceeds this revenue by the triangle bounded by the demand curve, the original supply curve, and the new quantity. This area — known as the “Harberger triangle” after the economist who analyzed it extensively — is equal to the excess burden or, equivalently, the deadweight loss of the tax. From this fact, the following hold true:

- The welfare loss from the tax is greater than the revenue that the tax yields.
- Therefore, the social benefit of the activity for which the government uses the revenue must be greater than the revenue used.
- Equivalently, the marginal social cost of each dollar raised by the government is, in general, greater than one dollar.

The excess burden of a tax increases with the square of the tax rate. Consequently, if the tax in figure 1.12 were 30 (7.5) instead of 15, the excess burden would have been four times greater (one-fourth as large), hence the importance of broadening the base and lowering the rate of taxation.

The ratio of the excess burden to tax revenue varies directly with the elasticity of demand and supply. Therefore, if either the elasticity of supply or the elasticity of demand is zero, the excess burden disappears. The most realistic example of completely inelastic supply is for agricultural markets. When a farmer harvests a perishable crop, he has the choice of selling it at whatever the market price is, or letting it spoil. It makes no sense for him to hold back his harvest from the market, so he will sell no matter what the price. In this case, his supply curve is perfectly inelastic. Figure 1.13 depicts this example. The tax falls completely on the farmer, and the price received by the farmer falls by the full amount of the tax (unlike in figure 1.12). A similar result holds with completely inelastic demand. In this case, however, the full incidence of the tax falls on the consumer.

A more likely example is one in which the elasticity of supply is infinite, rather than zero. This is the case in which, for instance, the product being taxed is traded internationally, and prices are determined in world markets. In this case, the producers in a particular country are price takers. These producers, however, will pass through any taxes that relate to a particular country or locale within a country because they have the option of selling in other markets. In this case (figure 1.14), the burden of the tax — and the excess burden — fall completely on the consumer. Moreover, because supply is completely elastic, the excess burden as a share of the total burden is higher than when supply is less elastic. Note that the ratio of excess burden to tax revenue is higher in figure 1.14 than in figure 1.12.
In a perfectly competitive market, in which suppliers are price takers, the most efficient tax system would set tax rates that are inversely proportional to the elasticity of demand:

\[
\frac{\tau_i}{\tau_j} = \frac{\eta_j}{\eta_i}
\]  
(3.A1),

where \( \tau \) is the tax rate on goods \( I \) and \( j \), and \( \eta \) is the elasticity of demand. Equation 3.A1 is known as the “Ramsey rule” after the economist who derived it in 1927. Although it is efficient, it may conflict with equity goals, if it results in high tax rates on necessities consumed by poor households.
The Contribution of Traditional Herbal Medicine Practitioners to Kenyan Health Care Delivery

Results from Community Health-Seeking Behavior Vignettes and a Traditional Herbal Medicine Practitioner Survey

John Lambert, Kenneth Leonard with Geoffrey Mungai, Elizabeth Omindi-Ogaja, Gladys Gatheru, Tabitha Mirangi, Jennifer Owara, Christopher H. Herbst, GNV Ramana, Christophe Lemiere

September 2011