Issue no. 65, 27 September 2013

In this issue:

Real-World Economics Review Blog

home   manifesto   join   donate   journals   conferences   national chapters   newsletters

Regression and causation: a critical examination of six econometrics textbooks

Bryant Chen and Judea Pearl   [University of California, Los Angeles, USA]

Abstract
This report surveys six influential econometric textbooks in terms of their mathematical treatment of causal concepts. It highlights conceptual and notational differences among the authors and points to areas where they deviate significantly from modern standards of causal analysis. We find that econononmetric textbooks vary from complete denial to partial acceptance of the causal content of econometric equations and, uniformly, fail to provide coherent mathematical notation that distinguishes causal from statistical concepts. This survey also provides a panoramic view of the state of causal thinking in econometric education which, to the best of our knowledge, has not been surveyed before.

1. Introduction

The traditional and most popular formal language used in econometrics is the structural equation model (SEM). While SEMs are not the only type of econometric model, they are the primary subject of each introductory econometrics textbook that we have encountered. An example of an SEM taken from (Stock and Watson, 2011, p. 3) is modeling the effect of cigarette taxes on smoking. In this case, smoking, Y, is the dependent variable, and cigarette taxes, X, is the independent variable. Assuming that the relationship between the variables is linear, the structural equation is written $Y = \beta X + \varepsilon$. Additionally, if X is statistically independent of ε, often called exogeneity, linear regression can be used to estimate the value of β, the "effect coefficient".

More formally, an SEM consists of one or more structural equations, generally written as $Y = X^t \beta + \varepsilon$ in the linear case, in which Y is considered to be the dependent or effect variable, $X = < X_1, X_2, \ldots, X_n >$ a vector of independent variables that cause Y, and $\beta = < \beta_1, \beta_2, \ldots, \beta_n >$ a vector of slope parameters such that $x^t \beta$ is the expected value of Y given that we intervene and set the value of X to x. Lastly, ε is an error term that represents all other direct causes of Y, accounting for the difference between $x^t \beta$ and the actual values of Y[1]. If the assumptions underlying the model are correct, the model is capable of answering all causally related queries, including questions of prospective and introspective counterfactuals[2]. For purposes of discussion, we will use the simplest case in which there is only one structural equation and one independent variable and refer to the structural equation as $Y = \beta X + \varepsilon$.

The foundations for structural equation modeling in economics were laid by Haavelmo in his paper, "The statistical implications of a system of simultaneous equations" (Haavelmo, 1943). To Haavelmo, the econometric model represented a series of hypothetical experiments. In his 1944 paper, "The Probabilistic Approach in Econometrics", he writes:

---

[1] A more precise definition of the SEM invokes counterfactuals and reads $X^t \beta + \varepsilon = Y(u)$, where $Y_X(u)$ is the counterfactual "the value that Y would take in unit u, had X been x" (see Simon and Rescher 1966, Balke and Pearl 1995, Heckman 2000, Pearl 2012b, and Appendix A).
[2] Prospective counterfactual queries are queries of the form, "What value would Y take if X were set to x?" Introspective counterfactual queries are queries of the form, "What would have been the value of Y if X had been set to x?"

> "What makes a piece of mathematical economics not only mathematics but
> also economics is, I believe, this: When we have set up a system of
> theoretical relationships and use economic names for the otherwise purely
> theoretical variables involved, we have in mind some actual experiment, or
> some design of an experiment, which we could at least imagine arranging, in
> order to measure those quantities in real economic life that we think might
> obey the laws imposed on their theoretical namesakes" (Haavelmo,
> 1944, p. 5).

Using a pair of non-recursive equations with randomized ε's, Haavelmo shows that $\boldsymbol{\beta x}$ in the equation $\boldsymbol{y = \beta x + \varepsilon}$ is not equal to the conditional expectation, $\boldsymbol{E\left[Y|x\right]}$, but rather to the expected value of $\boldsymbol{Y}$ given that we intervene and set the value of $\boldsymbol{X}$ to $\boldsymbol{x}$. This "intervention-based expectation" was later given the notation $\boldsymbol{E\left[Y|\mathbf{do}(x)\right]}$ in (Pearl, 1995)[3].

In the years following Haavelmo's 1944 paper, this interpretation has been questioned and misunderstood by many statisticians. When Arthur Goldberger explained that $\boldsymbol{\beta x}$ may be interpreted as the expected value of $\boldsymbol{Y}$ "if $\boldsymbol{x}$ were fixed," Nanny Wermuth replied that since $\boldsymbol{\beta x \neq E\left[Y|x\right]}$, "the parameters... cannot have the meaning Arthur Golberger claims" (Goldberger, 1992; Wermuth, 1992).

(Pearl, 2012b) summarizes the debate in the following way: For statisticians like Wermuth, structural coefficients have dubious meaning because they cannot be expressed in the language of statistics, while for economists like Goldberger, statistics has dubious substance if it excludes from its province all aspects of the data generating mechanism that do not show up in the joint probability distribution.

Econometric textbooks fall on all sides of this debate. Some explicitly ascribe causal meaning to the structural equation while others insist that it is nothing more than a compact representation of the joint probability distribution. Many fall somewhere in the middle – attempting to provide the econometric model with sufficient power to answer economic problems but hesitant to anger traditional statisticians with claims of causal meaning. The end result for many textbooks is that the meaning of the econometric model and its parameters are vague and at times contradictory.

We believe that the source of confusion surrounding econometric models stems from the lack of a precise mathematical language to express causal concepts. In the 1990s, progress in graphical models and the logic of counterfactuals led to the development of such a language (Pearl, 2000). Significant advances in causal analysis followed. For example, algorithms for the discovery of causal structure from purely observational data were developed (Verma and Pearl, 1990; Spirtes et al., 1993; Verma, 1993) and the problem of causal effect identifiability was effectively solved for non-parametric models (Pearl, 1995; Tian and Pearl, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006; Shpitser, 2008). These and other advances have had marked influence on several research communities (Glymour and Greenland, 2008; Morgan and Winship, 2007) including econometrics (Heckman, 2008; White and Chalak, 2009), but their benefits are still not fully utilized (Pearl, 2012b).The purpose of this report is to

---

[3] The expression $E\left[Y|\mathrm{do}(x)\right]$ can also be interpreted as the expected value of Y in an ideal randomized experiment for a subject assigned treatment $X = x$. Clearly, $E\left[Y|\mathrm{do}(x)\right]$ does not necessarily equal $E\left[Y|x\right]$. For example, the expected performance of an employee at an earning bracket of $X = x$ is different from the expected performance if management decides to set someone's earning to $X = x$. A simple recipe for computing $E\left[Y|\mathrm{do}(x)\right]$ for a given model is provided in Appendix A, which provides formal definitions of counterfactuals and their relations to structural equations and the do(x) operator.

examine the extent to which these and other advances in causal modeling have benefited education in econometrics. Remarkably, we find that not only have they failed to penetrate the field, but even basic causal concepts lack precise definitions and, as a result, continue to be confused with their statistical counterparts.

In this paper, we survey six econometrics textbooks in order to analyze their interpretation and usage of the econometric model and compare them to modern standards of causal analysis.

2. Criteria for evaluation

In evaluating textbooks, we ask the following questions: What does the author believe is the purpose of an econometric model? To which problems can it be applied? How does the author interpret the model parameters and the structural equation? Does the author consider βx to be equal to the expected value of Y given x, $E[Y|x]$, or the expected value of Y given that we intervene and set X to x, $E[Y|do(x)]$? Does the author make clear the assumptions necessary to answer the problems that econometrics is expected to solve?

To answer these questions, we formulated 11 evaluation criteria and grouped them under three categories. We also state the "ideal"[4] answers to these questions.

Applicability of econometric models

   1.   Does the author present example problems that require causal reasoning?
   2.   Does the author present example problems that require prediction alone?

A predictive problem is one of the form, "Given that we observe X to be x, what value can we expect Y to take?" Many econometrics textbooks begin with example problems that they expect econometric methods to solve. We use these examples to determine the author's view on the purpose and applicability of the econometric model. Since both predictive and causal problems are of interest to economists, both should be exemplified in econometrics textbooks.

Interpreting model parameters

   3.   Does the author state that each structural equation in the econometric model is meant to convey a causal relationship?
   4.   Does the author define β by the equality, $\beta X = E[Y|X]$?

Clearly, since the structural equation represents a causal relationship between *X* and *Y*, it is incorrect to *define β* by $\beta X = E[Y|X]$, though the equality may occasionally be satisfied.

   5.   Does the author define the error term as being the difference between $E[Y|X]$ and *Y* ?
   6.   Does the author interpret the error term as omitted variables that (together with *X*) determine *Y* ?
   7.   Does the author state that each structural equation in the econometric model is meant to capture a *ceteris paribus* or "everything else held fixed" relationship?

---

[4] By "ideal" we mean consistent with modern analysis, as expressed in articles dealing specifically with the causal interpretation of structural equation models (Heckman, 2008; Leamer, 2010; Nevo and Whinston, 2010; Keane, 2010; Pearl, 2012a).

The notion of *ceteris paribus* is sometimes used by economists and is closely tied to direct causation. If we hold all other variables fixed then any measured relationship between X and Y must be causal. When we write $Y = \beta X + \varepsilon$, where ε represents all other direct causes of Y, then β must capture a *ceteris paribus* and, therefore, causal relationship between X and Y. It is for this reason that we examined whether the author explicitly states that the structural equation captures a *ceteris paribus* relationship.

8. Does the author assume that exogeneity of X is inherent to the model?

Economists consider X to be exogenous in the equation $Y = \beta X + \varepsilon$ if X is independent of ε, where ε represents all factors that have influence on Y when X is fixed[5]. An example of exogeneity is an ideal randomized experiment. Subjects are randomly assigned to a treatment or control group, ensuring that X is distributed independently of all personal characteristics of the subject. As a result, X and ε are independent and X is exogenous. Clearly, if X is exogenous β can be estimated using linear regression.

However, if βX is incorrectly interpreted as $E[Y|X]$ and ε incorrectly defined as $Y - E[Y|X]$ (as is done in the text by Hill, Griffiths, and Lim) then ε will always be uncorrelated with X and the statement that X is uncorrelated with ε is vacuous.

Moreover, if all we care about is the conditional expectation then it does not matter whether confounders or other causal biases are present, as regression will allow proper estimation of the slope of the equation $E[Y|X] = \alpha X$ so long as the relationship between X and $E[Y|X]$ is linear. In contrast, forcing X to be exogenous (e.g. through a randomized experiment) will estimate the interventional expectation and not the conditional expectation, which are not necessarily equal.

While exogeneity allows for unbiased estimation of β, it should not be considered an implicit assumption of the model. β retains its causal interpretation as $\beta = \frac{\partial}{\partial X} E[Y|\mathbf{do}(X)]$ regardless of whether X and ε are correlated or not.

Moreover, exogeneity is a sufficient but not a necessary condition for identification. By requiring that exogeneity be a default assumption of the model, we limit its application to trivial and uninteresting problems, providing no motivation to tackle more realistic problems (say, through the use of instrumental variables).

Distinguishing $E[Y|X]$ and $E[Y|\mathbf{do}(X)]$

9. Does the author make clear the difference in the assumptions needed for answering causal as opposed to predictive problems?
10. Does the author use separate notation for $E[Y|\mathbf{do}(X)]$ and $E[Y|X]$?
11. Does the author use separate notation for the slope of the line associated with $E[Y|X]$ and that associated with $E[Y|\mathbf{do}(X)]$?

---

[5] From a causal analytic perspective, X is exogeneous if $E[Y|X] = E[Y|\mathrm{do}(X)]$ (Pearl, 2000). However, for purposes of this paper, we will use the aforementioned definition in which X is exogenous if it is independent of ε. Note that if X is independent of ε then $E[Y|X] = E[Y|\mathrm{do}(X)]$. The converse may not hold. For example, when ε is a vector of factors with cancelling influences on Y

Many books present both predictive and interventional problems as applications for econometric analysis. Not all of them discuss the distinction between them despite the fact that they require fundamentally different assumptions and, at times, a different methodology. At the core of this distinction is whether the model is meant to estimate $E[Y|X]$ or $E[Y|\text{do}(X)]$. Clearly, if $\beta X = E[Y|\text{do}(X)]$ is estimated (as opposed to $E[Y|X]$) when attempting to make predictions, the answer may be drastically wrong. Utilizing explicit notation for the interventional distributions is essential for avoiding such errors.

Remarkably, all of the econometrics textbooks surveyed refer to the structural equation as the "regression" equation. This is another source of confusion because "regression" is used to refer to the best-fit line. Using the same term to refer to both the structural and best-fit lines further increases the confusion between interventions and predictions.

Results

We surveyed the following textbooks:
- Greene, W. *Econometric Analysis*. Pearson Education, New Jersey. 7th edition, 2012.
- Hill, R., Griffiths, W., and Lim, G. *Principles of Econometrics*. John Wiley & Sons Inc. New York. 4th edition, 2011.
- Kennedy, P. *A Guide to Econometrics*. Blackwell Publishers, Oxford. 6th edition, 2008.
- Ruud, P. *An Introduction to Classical Econometric Theory*. Oxford University Press, Oxford. 1st edition, 2000.
- Stock, J., Watson, M. *Introduction to Econometrics*. Pearson Education, Massachusetts. 3rd edition, 2011.
- Wooldridge. *Introductory Econometrics: A Modern Approach*. South-Western College Pub. 4th edition, 2009.

These are six highly popular and frequently cited introductory econometrics textbooks. Our results are summarized in Table 1.

Table 1: Summary of survey results

	Greene	**Hill, Griffiths, Lim**	**Kennedy**	**Stock, Watson**	**Ruud**	**Wooldridge**	**Ideal**
1.	Yes ✓	Yes ✓	No	Yes ✓	No[6]	Yes ✓	Yes
2.	No	Yes ✓	No	Yes ✓	Yes ✓	No	Yes
3.	No[7]	No	No	Yes ✓	No	Yes ✓	Yes
4.	No[8]	Yes ✗	Yes ✗	Yes[9] ✗	No ✓	No ✓	No
5.	No ✓	Yes ✗	No ✓	No ✓	No ✓	No ✓	No
6.	Yes ✓	Yes[10] ✓	Yes ✓	Yes ✓	No[11]	Yes ✓	Yes
7.	Yes ✓	No	No	No[12] ✓	No	Yes ✓	Yes
8.	Yes	Yes	Yes	No ✓	No[13] ✓	Yes	No
9.	No	No ✗	No	Yes ✓	No	No	Yes
10.	No	No	No	No	No	No	Yes
11.	No	No	No	No	No	No	Yes

✓ denotes agreement with the ideal column,
✗ denotes a contradiction with another response in the same textbook

3.1 Greene (2012)

Greene writes, "The ultimate goal of the econometric model builder is often to uncover the deeper causal connections through elaborate structural, behavior models" (Greene, 2012, pp. 5-6). Consistent with this goal, Greene provides seven applications of econometric modeling as examples (*ibid.*, p. 3), each of which requires the estimation of causal effects. Among them are the effect of different policies on the economy, the effect of a voluntary training program in work environments, the effect of attending an elite college on future income, and the effect of smaller class sizes on student performance.

---

[6] Mentions that latent variable models can be used for policy analysis but does not provide examples.
[7] Discusses the regression equation as capturing the deterministic relationship between the independent and dependent variables and writes that "the ultimate goal of the econometric model builder is often to uncover the deeper causal connections through elaborate structural, behavior models". (Greene, 2012, p. 2).
[8] States "The **regression** of **y** on **X** is the conditional mean, $E\,[\mathbf{y}|\mathbf{X}]$, so that without [exogeneity], $\mathbf{X}\beta$ is *not* the conditional mean function" (Greene, 2012, p. 21). Also, "The unknown parameters of the stochastic relationship $y_i = \mathbf{x_i}\beta + \varepsilon_i$ are the objects of estimation... The **population regression** is $E\,[y_i|\mathbf{x_i}] = \mathbf{x_i}'\beta$ whereas our estimate of $E\,[y_i|\mathbf{x_i}]$ is denoted $\hat{y}_i = \mathbf{x_i}\mathbf{b}$." (Greene, 2012, p.26).
[9] States "The first part of Equation (4.5), $\beta_0 + \beta_1 X_i$, is the **population regression line** or the **pop-ulation regression function**. This is the relationship that holds between Y and X on average over the population. Thus, if you knew the value of X, according to this population regression line you would predict that the value of the dependent variable, Y, is $\beta_0 + \beta_1 X$" (Stock and Watson, 2011, p. 110).
[10] States that the error term is comprised of omitted factors that affect the independent variable, approximation errors that arise due to the functional specification being only an approximation, and any elements of "random behavior that may be present in each individual" (Hill et al., 2011).
[11] States that ε represents unobserved, explanatory random variables (Ruud, 2000, p. 493).
[12] While Stock and Watson do not discuss the relationship between β and *ceteris paribus* per se, they state that β represents a causal relationship and discuss its relationship to randomized experiments. As a result, they implicitly define βX as $E\,[Y\,|do(X)]$ and we denote agreement with the ideal response.
[13] Prior to introducing latent variable models, Ruud does not make any assumptions regarding exogeneity. He only writes, "if the mean of y conditional on X is $X\beta_0$, the OLS estimator is unbiased: $E[\hat{\beta}X] = \beta_0$" (Ruud, 2000, p. 173). After introducing the latent variable model as $y_n = \mathbf{x}_n'\beta_0 + \varepsilon_n$, he writes, "In each model that we describe, at least one of the explanatory variables in $\mathbf{x}_n$ is correlated with ε_n so that $E[\varepsilon_n|\mathbf{x}_n]$ is a function of $\mathbf{x}_n$ and, therefore, not zero. This in turn implies that $E[y_n|\mathbf{x}_n] \neq \mathbf{x}_n\beta_0$ and that the OLS fit of y_n to x_n will yield inconsistent estimates of β_0" (Ruud, 2000, p. 491).

Although Greene acknowledges the goal of economic modeling to be the establishment or estimation of "causal connections", he does not explicitly discuss the role of model param-eters in pursuing this goal and refrains from attributing causal interpetation to *β*. Instead, he relates econometric models to the conditional expectation, writing, "The model builder, thinking in terms of features of the conditional distribution, often gravitates to the expected value, focusing attention on $E[y|x]$..." (*ibid.*, p. 12). At the same time, Greene also suggests that *β* carries meaning beyond that of the conditional expectation, writing, "The regression of y on **X** is the conditional mean, $E[y|X]$, so that without [exogeneity], $X\beta$ is not the conditional mean function" (*ibid.*, p. 21). He does not, however, tell readers what *β* stands for, what it is used for, or why it justifies all the attention given to it in the book. Instead, he writes, "For modeling purposes, it will often prove useful to think in terms of 'autonomous variation.' One can conceive of movement of the independent variables outside the relationship defined by the model while movement of the dependent variable is considered in response to some independent or exogenous stimulus" (*ibid.*, p. 13). While this may be a legitimate way of thinking about causal effects, depriving "*β*" of its causal label creates the impression that economic models incorporate ill-defined parameters that require constant re-thinking to ascertain their interpretation[14].

Later, when discussing endogeneity and instrumental variables, Greene seems to suggest that a natural experiment and instrumental variable is needed to bestow causal meaning to *β*. He writes,

> "The technique of instrumental variables estimation has evolved as a mechanism for disentangling causal influences... when the instrument is an outcome of a 'natural experiment,' true exogeneity is claimed... On the basis of a natural experiment, the authors identify a cause-and-effect relationship that would have been viewed as beyond the reach of regression modeling under earlier paradigms" (*ibid.*, p. 252).

Here the reader wonders why the coefficient *β*, considered under endogeneity, would not deserve the title "cause and effect relationship" unless a good instrument is discovered by imaginative authors. Up to this point, Greene has made only passing references to the relationship between structural parameters (e.g., *β*), regression, and causality.

In section 19.6, "Evaluating Treatment Effects", however, Greene introduces potential outcomes and discusses causal effects explicitly (*ibid.*, p. 889); gone are the hesitation and ambiguities that marred the discussion of structural equations. Here, Rubin's notation for counterfactuals is introduced and Greene discusses the estimation of causal effects using regression, propensity score matching, and regression discontinuity (instrumental variables are mentioned in an earlier chapter). However, Greene provides no connections between treatment effects defined in this chapter and the structural equations that were the subject of discussion in the 18 earlier chapters. The impression is, in fact, created that the previous chapters were a waste of time for researchers aiming to estimate causal effects, which the book defines as, "The ultimate goal of the econometric model builder".

In section 19.6.1, "Regression Analysis of Treatment Effects", Greene presents the equation,

---

[14] In a personal correspondence (2012), Greene wrote, "The precise definition of effect of what on what is subject to interpretation and some ambiguity depending on the setting. I find that model coefficients are usually not the answer I seek, but instead are part of the correct answer. I'm not sure how to answer your query about exactly, precisely carved in stone, what *β* should be."

$earnings_i = x_i^t \beta + \delta C_i + \varepsilon_i$ and asks,

> "Does δ measure the value of a college education (assuming that the rest of the regression model is correctly specified)? The answer is no if the typical individual who chooses to go to college would have relatively high earnings whether or not he or she went to college..."

The answer is, in fact, YES[15]. The only way to interpret Greene's negative answer is to assume that the equation is regressional and that δ is simply the slope of the regression line. However, as mentioned above, Greene also suggests that "regression" parameters (*ibid.*, p. 21) are more than just slopes of regression lines. Indeed, this is the interpretation that is generally used throughout the textbook. This inconsistency is a major source of confusion to students attempting to understand the meaning of parameters like "β" or "δ". In summary, while Greene provides the most detailed account of potential outcomes and counterfactuals of all the authors surveyed, his failure to acknowledge the oneness of the potential outcomes and structural equation frameworks is likely to cause more confusion than clarity, especially in view of the current debate between two antagonistic and narrowly focused schools of econometric research (See Pearl 2009, p. 379-380).

3.2 Hill, Griffiths, and Lim (2011)

In the first chapter of the text by Hill, Griffiths, and Lim, the authors discuss the role of econometrics in aiding both prediction and policy making. On pp. 3-4, they present several problems as examples, some of which are causal and some of which are predictive:

- "A city council ponders the question of how much violent crime will be reduced if an additional million dollars is spent putting uniformed police on the street.
- "The owner of a local Pizza Hut must decide how much advertising space to purchase in the local newspaper, and thus must estimate the relationship between advertising and sales.
- "You must decide how much of your savings will go into a stock fund, and how much into a money market. This requires you to make predictions of the level of economic activity, the rate of inflation, and interest rates over your planning horizon (Hill et al., 2011)".

However, in explaining the meaning and usage of the econometric model, the text makes no mention of causal vocabulary and instead relies on statistical notions like conditional expectation. For example, on p. 43, they write, "the economic model summarizes what theory tells us about the relationship between [x] and the... $E(y|x)$" and the "simple regression function" of the model is defined as $E(y|x) = \beta_1 + \beta_2 x$ (*ibid.*, pp. 43) where β_1 is defined as $E(y|x = 0)$ and β_2 as $\frac{dE(y|x)}{dx}$. At no point is causality or *ceteris paribus* mentioned.

This interpretation leaves the econometric model unable to guide policy making and solve the aforementioned problems requiring causal inference. Indeed, these problems seem to be forgotten in chapter 2 when the econometric model is introduced and instead, we find only predictive examples: "An econometric analysis of the expenditure relationship can provide

---

[15] *δ,* in this structural equation, measures precisely the value of a college education, regardless of what sort of individuals choose to go to college. While the OLS estimation of *δ* will be biased, the meaning of *δ* remains none other but the "value of college education".

answers to some important questions, such as: If weekly income goes up by $20, how much will average weekly food expenditure rise? Or, could weekly food expenditures fall as income rises? How much would we predict the weekly expenditure on food to be for a household with an income of $800 per week?" (*ibid.*, p. 44).

At the same time, when discussing the assumptions inherent to the econometric model the text states that "the variable *x* is not random" (*ibid.*, p. 45) and explains this assumption using an example of a McDonald's owner "[setting] the price (*x*) and then [observing] the number of Big Macs sold (*y*) during the week. The following week the price could be changed, and again the data on sales collected." (*ibid.*, p. 46 - 47). Clearly, requiring that the data be generated by a process in which X is fixed by intervention suggests that $\beta_1 + \beta_2 x$ has meaning beyond that of the $E(y|x)$.

Later, the authors introduce the error term as $e = y - E(y|x)$ (*ibid.*, p. 46) and the regression equation is defined as $y = \beta_1 + \beta_2 x + e$. Using these definitions, they relax the assumption that *x* be "fixed" explaining that it is unnecessary so long as it is uncorrelated with the error term (*ibid.*, p. 402). Not only is the requirement that *e* be uncorrelated with X redundant when *e* is defined as the residual, $y - E(y|x)$, but relating it to the assumption that *x* is "not random" leaves readers in a state of total confusion regarding the meaning of *β*.

3.3 Kennedy (2008)

Kennedy introduces the structural model using an example where consumption, *C*, is the dependent variable, and income *Y* is the independent variable. He writes the structural equation as $C = f(Y) + \varepsilon$ or $C = \beta_1 + \beta_2 Y + \varepsilon_i$ in the linear case, where *ε* is a disturbance term, and adds, "Without the disturbance term the relationship is said to be *exact* or *deterministic*..." (Kennedy, 2008, p. 3). Kennedy then writes that "some econometricians prefer to define the relationship between *C* and *Y* discussed earlier as 'the mean of *C* conditional on *Y* is $f(Y)$,' written as $E(C|Y) = f(Y)$." This [says Kennedy] "spells out more explicitly what econometricians have in mind when using this specification" (*ibid.*, p. 9). This unfortunately is wrong; the conditional interpretation $E(C|Y) = f(Y)$ is precisely what econometricians *do not* have in mind in writing the structural equation $C = f(Y) + \varepsilon$.

Both Haavelmo (1943) and Goldberger (1992) have warned econometricians of the pitfalls lurking in this interpretation. Oddly, Kennedy is well aware of the difference between the two interpretations and writes: "The conditional expectation interpretation can cause some confusion" (*ibid.*), yet he fails to tell readers which of the two interpretations they should adopt and why the conditional interpretation does not capture "what econometricians have in mind when using this specification".

Kennedy later suggests that causality has no place in econometric modeling and all uses of the term "cause" should be replaced with "Granger-cause". He writes, "Granger developed a special definition of causality which econometricians use in place of the dictionary definition; strictly speaking, econometricians should say 'Granger-cause' in place of 'cause', but usually they do not" (*ibid.*, p. 63). As is well known, and as Granger repeatedly stated[16], "Granger causality" is a misnomer given to purely predictive notion that has nothing to do with causation. Thus, Kennedy views economic models to be used strictly in prediction tasks and not as guides to policy making. Unfortunately this contradicts a claim made later in

---

[16] Granger, in a personal communication with J. Pearl, Uppsala, 1991.

the book that econometric model *can* be used to simulate the effects of policy changes (*ibid.*, p. 343).

Like Hill, Griffiths, and Lim, on page 41, Kennedy writes that one of the assumptions of the "classical linear regression model" (CLR) is that "the observations on the independent variables... be fixed in repeated samples" (*ibid.*, p. 41). While it is not immediately clear whether "fixed in repeated samples" is meant to imply active intervention on the independent variable or merely "repeated at the same observed value of *x*", in a later chapter, Kennedy discusses when this assumption is violated and writes, "In many economic contexts the independent variables are themselves random (or stochastic) variables and thus could not have the same value in repeated samples" (*ibid.*, p. 137). He then writes that "the assumption of fixed regressors is made mainly for mathematical convenience... If the assumption is weakened to allow the explanatory variables to be stochastic but to be distributed independently of the error term, all the desirable properties of the OLS estimator are maintained..." (*ibid.*). From this the reader may conclude, albeit indirectly, that "fixing" is related to exogeneity, that *x* should be fixed by intervention, and that the structural equation does capture a causal relationship, contrary to Kennedy's earlier suggestion that causality has no place in econometrics.

3.4   Ruud (2000)

Rather than treating an econometric model as representing an economic theory and testing it against data, Ruud focuses almost entirely on regression techniques. To Ruud, the regression line, as well as the mean, median, mode, and standard deviation, is a worthy descriptor of the dataset. Much of the textbook is devoted to deriving statistical properties of OLS regression. The exogeneity assumption and the equation, $y = \beta X + \varepsilon$, are introduced later in a chapter on instrumental variables as a latent variable model. Ruud mentions that latent variable models "play a key role in the economist's search for structure", "[assist] in the marriage of theoretical and empirical modeling", and can be used for policy analysis due to their "invariant features" (Ruud, 2000, p. 616) but does not discuss the way in which they can be used to accomplish the aforementioned goals and solve causal problems. Instead, he spends considerable effort explaining the statistics of latent variable models without discussing their relationship to structure and causality. In fact, causality is not discussed at all in the textbook beyond a passing mention that the causal effect and the conditional expectation are not the same. While this statistical approach is logically consistent, it leaves students unequipped to tackle causal problems.

3.5   Stock and Watson (2011)

The textbook by Stock and Watson explicitly discusses policy questions (hence cause-effect relations) in the econometric model. In the first chapter, they write that the "book examines several quantitative questions taken from current issues in economics. Four of these questions concern education policy, racial bias in mortgage lending, cigarette consumption, and macro-economic forecasting..." (Stock and Watson, 2011, p. 1). The authors acknowledge that three of these problems "concern causal effects" while "the fourth – forecasting inflation – does not" (*ibid.*, p. 9). Of the six textbooks surveyed, this text is the only one to address the difference in assumptions needed for causal versus predictive inference. They write, "when regression models are used for forecasting, concerns about external validity are very important, but concerns about unbiased estimation of causal effects are not" (*ibid.*, p. 327).

In addition to discussing the difference in predictive versus causal inference, the textbook also notes that coefficients of confounding variables added to regression equations for purposes of adjustment cannot be given a causal interpretation (*ibid.*, p. 232). At one point, the text even provides separate notation for such coefficients, labeling them *δ* as opposed to *β* (*ibid.*, p. 250). It would have been helpful to make this notational distinction consistent throughout the book, to clearly separate causal from regression coefficients, and to refrain from referring to structural equations as "regression".

The textbook also introduces the potential outcome framework to explain randomization and heterogeneous causal effects (*ibid.*, pp. 498-99). However, the relationship between potential outcomes and the structural equation is often obscured. For example, the authors write: "The potential outcomes framework, combined with a constant treatment effect, implies the regression model in $[Y_i = \beta_1 + \beta_2 X_i + u_i, i = 1, \ldots, n]$" (*ibid.*, p. 514). The sentence is misleading on two counts. First, the equation is not regressional but structural. Second, the structural equation is not a consequence of the potential outcomes framework but the other way around; the equation provides the scientific basis from which the potential outcomes framework draws its legitimacy (Pearl, 2000; Heckman, 2005; Pearl, 2012b)[17]. Nevertheless, this and the textbook by Greene are the only two surveyed that introduce the potential outcomes notation, which is important for defining counterfactual questions such as the effect of treatment on the treated and indirect effects.

Additionally, in contrast to the previous textbooks, this text recognizes and discusses the causal nature of the exogeneity condition. They write, "The random assignment typically is done using a computer program that uses no information about the subject, ensuring that *X* is distributed independently of all personal characteristics of the subject. Random assignment makes *X* and *u* independent, which in turn implies that the conditional mean of *u* given *X* is zero. In observational data, *X* is not randomly assigned in an experiment. Instead, the best that can be hoped for is that *X* is *as if* randomly assigned, in the precise sense that $E(u_i|X_i) = 0$[18]" (Stock and Watson, 2011, p. 123).

While the textbook provides a clearer explanation of the difference between causal and statistical concepts than the other textbooks surveyed, it still falls victim to prevailing habits in the economics literature. For example, after presenting an example in which *β* measures a causal effect, the text turns around and suggests that $E[Y|x] = \beta x$ (*ibid.*, pp. 108-10)[19].

More seriously, the authors state that "the slope of the line relating *X* and *Y* is an unknown characteristic of the population joint distribution of *X* and *Y*" (*ibid.*, p. 107). While this is probably a semantic slip, it risks luring readers back into the dark era when economic models were thought to represent joint distributions (see "Econometric Models", Wikipedia, August 2012). The structural slope, *β*, is NOT a characteristic of the "joint distribution of *X* and *Y*"; it is a characteristic of the data generating process but has no counterpart in the joint distribution.

---

[17] Appendix 1 of (Pearl, 2012b) provides explicit discussion of this point and demonstrates how the experimental and quasi-experimental ramification of the potential outcome framework are derived from ordinary structural equations. See also Appendix A.

[18] This is not strictly true; one can do better than hope for an *as if* miracle. Identification techniques are available for models in which *X* is far from satisfying $E(u_i|x_i) = 0$ (Pearl, 2000).

[19] In a personal correspondence James Stock acknowledged this correctable oversight.

3.6 Wooldridge (2009)

The textbook by Wooldridge also explicitly ascribes causal meaning to the econometric model. He writes, "In most tests of econometric theory, and certainly for evaluating public policy, the economist's goal is to infer that one variable (such as education) has a causal effect on another variable (such as worker productivity)" (Wooldridge, 2009, p. 12). In contrast to Stock and Watson, who define causality in relation to a randomized experiment (Stock and Watson, 2011, p. 6), Wooldridge emphasizes the concept of *ceteris paribus*. He writes, "You probably remember from introductory economics that most economic questions are ceteris paribus by nature. For example, in analyzing consumer demand, we are interested in knowing the effect of changing the price of a good on its quantity demanded, while holding all other factors fixed. If other factors are not held fixed, then we cannot know the causal effect of a price change on quantity demanded[20]." (Wooldridge, 2009, p. 12).

Wooldridge is also more careful when interpreting the parameter, *β*. Rather than using the conditional expectation of *Y* given *X*, he writes that *β* is "the slope parameter in the relationship between *y* and *x* holding the other factors in *u* fixed" (*ibid.*, p. 23), where *u* represents the error term.

While Wooldridge provided a strong and generally consistent account of causality, he did not provide explicit notation for intervention thus letting the definitions of beta and epsilon rest entirely on verbal description. While this may be adequate for linear models, it prevents one from extending causal analysis to nonparametric models.

4. Discussion and recommendations

4.1 Potential points of improvement

Five of the six authors surveyed claim that exogeneity of *X* is necessary for unbiased estimation of *β* using linear regression, indirectly implying that *β* has meaning beyond that of a regression coefficient. Only two of them explicitly ascribe causal meaning to the model.

We believe that making clear the difference between the conditional expectation, $E[Y|X]$, and the interventional expectation, $E[Y|do(X)]$, will do much to clarify the meaning of the econometric model and help prevent both students and economists from confusing the two.

It is common for textbook authors to equate the conditional expectation with *βX* even when it is clear that the author considers βX to be $E[Y|do(X)]$ rather than $E[Y|X]$. Of the five authors that claim exogeneity is necessary for unbiased estimation of *β* using linear regression, three also claim that $E[Y|X] = \beta X$. Kennedy admitted (personal correspondence, 2001) that he was careless in the 1998 edition and had intended for the statement to be applicable only when *X* is exogenous. However, $E[Y|X]$ is precisely *not* what economists have in mind when authoring an econometric model. This fact becomes even more evident when adjusting for a confounder or using instrumental variables in cases where βX is not equal to $E[Y|X]$. Economists developed these techniques precisely because in their minds *β* represents the causal effect of *X* on *Y*, not some property of the joint distribution.

---

[20] Again, this is not strictly true. There are many techniques that allow unbiased estimation of causal effects even when other factors are not held fixed (Pearl, 2000).

We have limited our comparison criteria to features that hinder basic understanding of the meaning of structural economic models – the absence of distinct causal notation. Lines 10-11 of Table 1 represent this deficiency, which is common to all six textbooks. In addition to the confusion it causes, it also results in technical limitations including, for example, inability to extend causal analysis to nonparametric models and forgoing the benefits of Marschak's Maxim[21] (Heckman, 2010).

Another weakness that runs across all books surveyed is the absence of graphical models to assist in both understanding the causal content of the equations and performing necessary inferential functions that are not easily performed algebraically. Introducing simple graphical tools would enable econometric students to recognize the testable implications of a system of equations; locate instruments in such systems; decide if two systems are equivalent; if causal effects are identifiable; if two counterfactuals are independent given another; and whether a set of measurements will reduce bias; and, most importantly, read and scrutinize the causal and counterfactual assumptions that such systems convey. The power of these tools is demonstrated in (Pearl, 2012a) and we hope to see them introduced in next-generation econometric textbooks.

We fully recognize, though, that authors in economics are reluctant to adopt, or even examine the power of graphical techniques, which generations of economists have dismissed (under the rubric of "path analysis") as "informal", "heuristic", or "mnemonic" (Epstein, 1987; Pearl, 2009, p. 138-139). For example, only a handful of economists have come to realize that graphical models have laid to rest the problem of identification in the entire class of "nonadditive, nonseparable triangular models"[22], for both discrete and continuous variables. We therefore offer our recomendations (below) in terms of essential problem-solving skills without advocating a specific notation or technique.

4.2 What an ideal textbook should contain

First and foremost, an ideal textbook in econometrics should eradicate the century-old con-fusion between regression and structural equations. Structural and regression parameters should consistently be given distinct notation, for example, β_s vs. α_r . The term "regression" should not be used when referring to structural equations. The assumptions behind each structural equation should be made explicit and contrasted with those that underlie regression equations. Policy evaluation examples should demonstrate the proper use of structural versus regression parameters in achieving the target estimates.

Additionally, students should acquire the following tools and abilities:
1. Ability to correctly classify problems, assumptions and claims into two distinct categories: causal vs. associational.
2. Ability to take a given policy question, and articulate mathematically both the target quantity to be estimated, and the assumptions that one is prepared to make (and defend) to facilitate a solution.
3. Ability to determine, in simple models, whether control for covariates is needed for estimating the target quantities, what covariates need be controlled, what the resulting estimand is, and how it can be estimated using the observed data.

---

[21] Marschak Maxim refers to Jacob Marschak's (1953) observation that many policy questions do not require the estimation of each and every parameter in the system – a combination of parameters is all that is necessary – and that it is often possible to identify the desired combination without identifying the individual components.
[22] We are using the nomenclature of (Matzkin, 2007). By "handful" we include (White and Chalak, 2009) and (Hoover, 2009). The graphical solution can be found in (Shpitser and Pearl, 2006, 2008).

4. Ability to take a simple model, determine whether it has statistically testable implications, then apply data to test the model for misspecification.

5. Finally, students should be aware of nonparametric extensions to traditional linear structural equations. In particular, they should be able to solve problems of identification and misspecification in simple nonparametric models, where no commitment is made to the form of the equations or to the distribution of the disturbances.

Examples of specific problems requiring these abilities are illustrated in (Pearl, 2012b, Section 3.2).

5. Conclusion

The surveyed econometrics textbooks range from acknowledging the causal content of the SEM (e.g.Wooldridge, Stock and Watson) to insisting that it is nothing more than a compact representation of a joint distribution (e.g. Ruud). The rest fall somewhere in the middle, attempting to provide the model with power to answer economic questions but unwilling to accept its causal nature; the result is ambiguity and confusion. Nowhere is this more evident than in the text by Hill, Griffiths, and Lim in which definitions of the model parameters conflict with stated assumptions of the model. Other textbooks (e.g. Greene) are more careful about avoiding contradictions but their refusal to acknowledge the causal content of the model results in ambiguous descriptions like "autonomous variation". Finally, even textbooks that acknowledge the role of causality in econometrics fail to provide coherent mathematical notation for causal expressions, luring them into occasional pitfalls (e.g. equating β with a regression coefficient or some other property of the joint distribution of X and Y) and preventing them from presenting the full power of structural equation models.

The introduction of graphical models and distinct causal notation into elementary econometric textbooks has the potential of revitalizing economics education and bringing next generation economists to par with modern methodologies of modeling and inference.

Appendix A

This appendix provides formal definitions of interventions and counterfactuals as they have emerged from Haavelmo's interpretation of structural equations. For a more detailed account, including examples of policy-related tasks, see (Pearl, 2012b).

Key to this interpretation is a procedure for reading counterfactual information in a system of economic equations, formulated as follows:

Definition 1 (unit-level counterfactuals) (Pearl, 2000, p. 98)**.**
Let M be a fully specified structural model and X and Y two arbitrary sets of variables in M. Let M_X be a modified version of M, with the equation(s) determining X replaced by the equation(s) X = x. Denote the solution for Y in the modified model by the symbol $Y_{M_x}(u)$, where u stands for the values that the exogenous variables take for a given individual (or unit) in the population. The counterfactual $Y_X(u)$ (Read: "The value of Y in unit u, had X been x") is defined by:

$$Y_x(u) \equiv Y_{M_x}(u) \tag{A.1}$$

In words: The counterfactual $Y_x(u)$ in model M is defined by the solution for Y in the modified submodel M_X, with the exogenous variables held at U = u.

For example, consider the model depicted in Figure 1(a), which stands for the structural equations:

$$Y = f_Y(X, Z, U_Y)$$
$$X = f_X(Z, U_X)$$
$$Z = f_Z(U_Z)$$

Here, f_X, f_Y, f_Z are arbitrary functions and U_X, U_Y, U_Z are arbitrarily distributed omitted factors. The modified model M_X consists of the equations below and and is depicted in Figure 1b.

$$Y = f_Y(X, Z, U_Y)$$
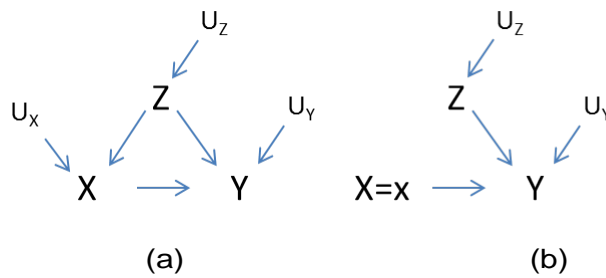$$X = x$$
$$Z = f_Z(U_Z)$$

(a)   (b)

Figure 1

The counterfactual $Y_x(u)$ at unit $u = (u_x, u_y, u_z)$ would take the value
$Y_x(u) = f_Y(x, f_Z(u_z), u_Y)$, which can be computed from the model. When u is unknown, the counterfactual becomes a random variable, written $Y_x(u) = f_Y(x, Z, U_Y)$ with x treated as constant, and Z and U_Y random variables governed by the original model.

Clearly, the distribution $P(Y_X = y)$ depends on both the distribution of the exogenous variables $P(Y_X, U_Y, Y_Z)$ and on the functions f_X, f_Y, f_Z . In the linear case, however, the expectation $E[Y_X]$ is rather simple.

Writing:

$$Y = aX + bZ + U_Y$$
$$X = cZ + U_X$$
$$Z = U_Z$$

gives

$$Y_x = ax + bZ + U_Y$$

and

$$E(Y_x) = ax + bE(Z)$$

Remarkably, the average effect of an intervention can be predicted without making any commitment to functional or distributional form. This can be seen by defining an intervention operator do(*x*) as follows:

$$P(Y = y | \text{do}(x)) \equiv P(Y_x = y) \equiv P_{M_x}(Y = y) \qquad \text{(A.2)}$$

In words, the distribution of *Y* under the intervention do(*X* = *x*) is equal to the distribution of *Y* in the modified model M_X, in which the dependence of *Z* on *X* is disabled (as shown in Figure 1b).

Accordingly, we can use M_X to define average causal effects:

Definition 2 (Average causal effect).

The average causal effect of X on Y, denoted by $E[Y|\text{do}(x)]$ is defined by:

$$E[Y|\text{do}(x)] \equiv E[Y_x] = E[Y_{M_x}] \qquad \text{(A.3)}$$

Note that Definition 2 encodes the effect of interventions not in terms of the model's parameters but in the form of a procedure that modifies the structure of the model. It thus liberates economic analysis from its dependence on parameteric representations and permits a totally non-parametric calculus of causes and counterfactuals that makes the connection between assumptions and conclusions explicit and transparent.

If we further assume that the exogenous variables (U_X, U_Y, U_Z) are mutually independent (but arbitrarily distributed) we can write down the post-intervention distribution immediately, by comparing the graph of Figure 1b to that of Figure 1a. If the pre-intervention joint probability distribution is factored into (using the chain rule):

$$P(x, y, z) = P(z)P(x|z)P(y|x, z)$$
(A.4)

the post-intervention distribution must have the factor *P* (*x*|*z*) removed, to reflect the missing arrow in Figure 1b. This yields:

$$P(x, y, z | \text{do}(X = x_0)) = \begin{cases} P(z)P(y|x, z) & \text{if } x = x_0 \\ 0 & \text{if } x \neq x_0 \end{cases}$$

In particular, for the outcome variable *Y* we have $P(y|\text{do}(x)) = \sum_z P(z)P(y|x, z)$, which reflects the operation commonly known as "adjusting for *Z*" or "controlling for *Z*". Likewise, we have

$$E(Y|\text{do}(x)) = \sum_z P(z)E(Y|x, z),$$

which can be estimated by regression using the pre-intervention data.

In the simple model of Figure 1a the selection of *Z* for adjustment was natural, since *Z* is a confounder that causes both *X* and *Y*. In general, the selection of appropriate sets for adjustment is not a trivial task; it can be accomplished nevertheless by a simple graphical procedure (called "backdoor") once we specify the graph structure (Pearl, 2009, p. 79).

Equation A.1 constitutes a bridge between structural equation models and the potential outcome framework advanced by (Neyman, 1923) and (Rubin, 1974), which takes the controlled randomized experiment as its guiding paradigm but encounters difficulties articulating modeling assumptions. Whereas structural models encode causal assumptions in the form of functional relationships among realizable economic variables, the potential outcome framework requires those same assumptions to be encoded as conditional independencies among counterfactual variables, an intractible cognitive task.

Appendix B

This appendix, which provides supporting quotes for Table 1, has been omitted from the journal version of this survey and can be found in the full version at <http://ftp.cs.ucla. edu/pub/stat_ser/r395.pdf>.

Acknowledgments

References

Balke, A. and Pearl, J. (1995). Counterfactuals and policy analysis in structural mod- els. In *Uncertainty in Artificial Intelligence 11* (P. Besnard and S. Hanks, eds.). Morgan Kaufmann, San Francisco, 11–18.

Epstein, R. (1987). *A History of Econometrics.* Elsevier Science, New York.

Glymour, M. and Greenland, S. (2008). Causal diagrams. In *Modern Epidemiology* (K. Rothman, S. Greenland and T. Lash, eds.), 3rd ed. Lippincott Williams & Wilkins, Philadelphia, PA, 183–209.

Goldberger, A. (1992). Models of substance; comment on N. Wermuth, 'On block- recursive linear regression equations'. *Brazilian Journal of Probability and Statistics* **6** 1–56.

Greene, W. (2012). *Econometric Analysis.* 7th ed. Pearson Education, New Jersey.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11** 1–12. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Founda- tions of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

Haavelmo, T. (1944). The probability approach in econometrics (1944)*. Supplement to *Econometrica* 12 12–17, 26–31, 33–39. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, New York, 440–453, 1995.

Heckman, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* **115** 45–97.

Heckman, J. (2005). The scientific model of causality. *Sociological Methodology* **35** 1–97.

Heckman, J. (2008). Econometric causality. *International Statistical Review* **76** 1–27.

Heckman, J. (2010). Building bridges between structural and program evaluation ap- proaches to evaluating policy. *Journal of Economic Literature* **48** 356–398.

Hill, R., Griffiths, W. and Lim, G. (2011). *Principles of Econometrics*. 4th ed. John Wiley & Sons Inc., New York.

Hoover, K. (2009). Counterfactuals and causal structure. *Available at SSRN 1477531* .

Huang, Y. and Valtorta, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 217–224.

Keane, M. P. (2010). A structural perspective on the experimentalist school. *Journal of Economic Perspectives* **24** 47–58.

Kennedy, P. (2008). *A Guide to Econometrics*. Sixth ed. MIT Press, Cambridge, MA.

Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspec- tives* **24** 31–46.

Matzkin, R. (2007). Handbook of econometrics. Elsevier, Amsterdam.

Morgan, S. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. Cambridge University Press, New York, NY.

Nevo, A. and Whinston, M. D. (2010). Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives* **24** 69–82.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science* 5 465–480.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* 82 669–710.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2012a). The causal foundations of structural equation modeling. In *Handbook of Structural Equation Modeling* (R. H. Hoyle, ed.). Guilford Press, New York. In press.

Pearl, J. (2012b). Trygve Haavelmo and the emergence of causal calculus. Tech. Rep. R- 391, Department of Computer Science, University of California, Los Angeles, CA. Written for *Econometric Theory*, special issue on Haavelmo Centennial.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.

Ruud, P. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press, Oxford.

Shpitser, I. (2008). *Complete Identification Methods for Causal Inference*. Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.

Shpitser, I. and Pearl, J. (2006). Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. Richardson, eds.). AUAI Press, Corvallis, OR, 437–444.

Shpitser, I. and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* **9** 1941–1979.

Simon, H. and Rescher, N. (1966). Cause and counterfactual. *Philosophy and Science* **33** 323–340.
Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction, and Search*. Springer-Verlag, New York.

Stock, J. and Watson, M. (2011). *Introduction to Econometrics*. 3rd ed. Addison-Wesley, New York.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Pro- ceedings of the Eighteenth National Conference on Artificial Intelligence.* AAAI Press/The MIT Press, Menlo Park, CA, 567–573.

Verma, T. (1993). Graphical aspects of causal models. Tech. Rep. R-191, UCLA, Computer Science Department.

Verma, T. and Pearl, J. (1990). Equivalence and synthesis of causal models. In *Pro- ceedings of the Sixth Conference on Uncertainty in Artificial Intelligence.* Cambridge, MA. Also in P. Bonissone, M. Henrion, L.N. Kanal and J.F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence 6*, Elsevier Science Publishers, B.V., 255–268, 1991.

Wermuth, N. (1992). On block-recursive regression equations. *Brazilian Journal of Probability and Statistics* (with discussion) **6** 1–56.

White, H. and Chalak, K. (2009). Settable systems: An extension of pearl's causal model with optimization, equilibrium and learning. *Journal of Machine Learning Research* **10** 1759–1799.

Wooldridge, J. (2009)   Should instrumental variables be used as matching variables? Tech. Rep.<https://www.msu.edu/~ec/faculty/wooldridge/current%20research/treat1r6.pdf>   Michigan State University, MI

Author contacts:
   Bryant Chen: bryantc@cs.ucla.edu
   Judea Pearl: judea@cs.ucla.edu

_____

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

Diagrammatic economics

A review of Blaug, M. and Lloyd, P. eds. *Famous Figures and Diagrams in Economics*, Cheltenham, U.K.: Edward Elgar, 2010, pp. xvii, 468.

John Pullen   [University of New England, New South Wales, Australia]

Introduction

This book is a remarkable collection of figures and diagrams that have been prominent in the history of economics. In 58 chapters contributed by over 50 authors, it reproduces with expert commentary 58 named diagrams, dating from as early as 1758 to as recently as 1996. A further 89 diagrams have been used in commentaries on the 58 named diagrams. A splendid 23 page introduction by the editors provides an account of the origins of the diagrams and their places in the history of economics, together with a useful chronology of their discoverers and an appraisal of the contributions. The collection could be described as a celebration of over 200 years of geometrical skill and achievement in economics.

Part I: The purpose of diagrams

For those who love diagrams for their own sake, and who think that diagrams have made a significant contribution to the progress of economics and the progress of human welfare, this book is exciting and essential reading. But while acknowledging and admiring the scholarly expertise of the contributors and the effort of the editors in organising such a massive undertaking, its publication might be an opportune occasion to reflect on the purpose and achievements of diagrammatic economics and to ask whether these diagrams have made any contributions to economic thought that would not have been made just as effectively by a non-diagrammatic use of words, or whether they are merely gimmicks.

Most of the contributions to this Picture Book of Economics have certain features in common. The contributors have reproduced and explained the diagrams, but with possibly a few exceptions seem to have assumed that the diagrams are self-justifying and have made worthwhile contributions to the science of economics. There is a general absence of rigorous critique of the diagrams and few attempts to assess their practical usefulness. The contributors were probably not requested to do so and therefore could not reasonably be criticised for not doing so. Their collective achievement in presenting and explaining the diagrams is already an original contribution to the literature. The tenor of the book is therefore understandably uncritical or a-critical, but as in most areas of economics there are dissident voices which would like to be juxtaposed to this assemblage of praise for the diagrammatizers of economics.

The introduction makes a threefold claim: diagrams "have played a central role in the development of economic theory"; they "have been a major vehicle of discovery of economic concepts and propositions"; and "have excelled as an expository device". Such statements are sure to be condoned by all diagrammatizers, but also sure to infuriate all who abhor diagrams and who see diagrams as being largely responsible for the current distressed state of economics. We are told "figures are part of the basic textbook of modern economics", and assured that the diagrams selected for inclusion are commonly regarded as essential knowledge and "cover a large part of mainstream economic theory and analysis". Such

statements are presumably intended to be taken as an endorsement and justification of diagrammatic economics. An alternative view is that they should be interpreted as an indictment of what currently passes for mainstream economics and is a cause for regret rather than congratulations.

Although the editorial introduction claims that figures and diagrams have been used in economic theory "as a device to discover economic results" and "to prove some results", it somewhat undermines this claim when it adds that "it is not easy to uncover the role of figures and diagrams in the discovery of results", and when it states that although Alfred Marshall was an advocate and frequent user of diagrams, he warned that "graphical illustrations are not proofs. They are merely pictures corresponding very roughly to the main conditions of certain problems". The credibility of diagrammatical economics is further undermined when we read that the "use of geometry as a device for discovery and proof has declined in recent decades." Opponents of the diagrammatization of economics will no doubt hope that the decline will continue and that this volume, rather than being a celebration of diagrams past, will be a shrine of remembrance, or a mausoleum, or an addition to the already overstocked museum of economic antiquities and misdirected efforts.

The claim that diagrams "have been a major vehicle of discovery of economic concepts and propositions" is debatable. To be believable it would need to be substantiated by particular examples. Concepts and propositions are conceived and can lead to diagrams, but is the reverse likely? Has it ever happened, could it ever happen, that the diagram precedes the concept, logically or temporally? We are told that the Laffer Curve was first drawn on a restaurant napkin. Was it only after doodling it on the napkin and because of the doodling that Laffer conceived the idea that beyond a certain point further increases in the tax rate will reduce the total tax revenue? In how many of the 58 diagrams in this collection has the diagram being "a major vehicle of discovery" of the concept? Is there compelling evidence that any one of the diagrams preceded and caused its relevant concept?

The unspoken assumption underlying most if not all of the contributions to this collection seems to be that if it is a diagram, it is good. The book's aim is to show the role of diagrams "in the development of economic theory", but does not consider the possibility that "development" does not necessarily mean beneficial change. Dissenting critics would want to argue that the "development" of diagrammatic economics has been a regrettable digression in the history of economics and has retarded rather than advanced the science of economics.

Even more debatable is the claim that diagrams have a valuable pedagogic function and "have excelled and continued to excel" as an expository device. The editors ask rhetorically: "What teacher of economic theory has not seen the dawn of understanding come over students when, failing to understand the exposition of some complex model in algebra or calculus, they are presented with a simple illustration?" If some teachers economics have witnessed such an awakening, they are indeed fortunate, especially by contrast with other teachers who, when they present a diagram, encounter not the dawn of understanding but the twilight of unknowing. If this magical awakening of the spirit does occur, does it justify the use of diagrams, or does it merely prove that the diagram is less obscurantist than the algebra and calculus? For every student whose awakening dawn comes through diagrams there is another for whom diagrams are a crepuscular confusion. If the diagrammatic exposition is clearer than the verbal exposition, does this mean that diagrams are better than words, or could it merely be a reflection of the inadequacy of our ability as lecturers to express ourselves clearly in words? Do we resort to diagrams because of our verbal incompetence?

It is claimed that diagrams are "much more than illustrations" and that "in many areas of economics the way in which economists understand economic concepts and propositions is through figures and diagrams." But anti-diagrammatizers would argue that if it cannot be explained in words, it cannot be explained by diagrams, and if it can be explained in words, there is no need for diagrams. A possible defence of diagrams might be that they complement the words, and enhance the understanding; but a counter-argument is that diagrams in economics are beginning to supersede words. The question has to be asked: Would economics today be any worse off if economic diagrams had never been invented? Can we point to any diagrams in this collection that have materially advanced the cause of human welfare, apart from providing for the personal satisfaction and income of those who teach them? Are diagrams science-enhancing or career-enhancing? Has there been any significant real-world economic policy that would not have occurred were it not for an associated diagram?

There is no doubt that as intellectual exercises diagrams are excellent, but it could be argued that as aids to a better understanding of how economies work, they are about as useful as Sudoku; and that the profession has become self-hypnotized by diagrams, with its collective attention fixated on the captivating displays of lines and curves.

Nor should the opportunity cost of diagrammatic economics be ignored. If it occupies a large part of mainstream economics, what areas are being crowded out of the economics curriculum and neglected?

There is one undeniably useful function of diagrams in economics, and that is, as prosthetic devices for maintaining law and order in the classroom. All economics lecturers, even those of us with outstanding oratorical skills and charismatic personalities, know that when the attention of a class of 1000 restive students of microeconomic theory begins to wane, the only way to regain it is to show a diagram. For most lecturers, the idea of trying to conduct a 50-minute lecture without any diagrams is unthinkable; the concept of a non-diagrammatical economics is a self-contradiction. The use of water cannon is no longer socially acceptable as a means of crowd control in the lecture theatre, but diagrams are just as effective.

Another associated and undeniably useful function of diagrams, another reason for their creation, and another reason for the reluctance of lecturers to abandon them, is the ease with which they can be assessed in assignments and examinations. An experienced eye can judge and annotate a diagram in seconds, compared with the agony of wading through a 1000 word essay. The permanence of some diagrams adds further to their charm. Once the profession has accepted a standard presentation of the diagram, it becomes immutable and can be wheeled out year after year to each successive cohort of students without revision and without any adjustment of overheads, slides and power points. It is hard to imagine the cries of shock and horror if diagrams in economics were abolished. They are the bread and butter of the academic economist, the lecturer's friend and helper; and life without them would be far less comfortable.

While recognizing and commending the undoubted skill that goes into the creation and presentation of diagrams, is there are not a tendency in some textbooks to go beyond admiration into adulation, veneration, and adoration, with some diagrams being extolled and held aloft as quasi-religious icons, as uplifting symbols to inspire faithful devotees and convert infidels? Christianity has its cross, Islam has its crescent, Communist Russia has its

hammer and sickle, Nazi Germany has its swastika. For some members of the sect of economists known as classical Marshallians, the Marshallian cross becomes a talismanic symbol of their faith, a banner which sums up all they hold dear, and under which they serve and fight the good fight.

In reading the 58 chapters of this book, one has the impression that, at least in some cases, the presenters are more enamoured of the diagrams than the concepts behind them. Instead of merely being the medium by which the concept is conveyed, the diagram has become the message. It is given more attention than the idea; and for some diagram devotees is more interesting and more important than the idea. Were it not for their pictorial representation, some of the ideas would have been relatively insignificant or trivial. Arguments about the diagrams can generate more discussion and debate than arguments about the concepts, and become substitutes for arguments about the real world.

In the diagrammatic way of thinking, the diagram takes causal priority over the reality. We forget that the diagram is merely a representation of reality, not reality itself. The diagram usurps the concept and the reality. For those brought up in the diagrammatic world, arguing by diagrams seems to be more convincing than arguing by words; they appear to assume that no argument can be conclusively settled without recourse to diagrams.
.
The urge to mathematize and diagrammatize is extremely strong, and the self-satisfaction from successfully doing so is very rewarding, so much so that the drawing of diagrams becomes an end in itself, irrespective of whether the process is at all conducive to the progress of the human condition. Diagrams are presumably conceived as a means towards the end of better understanding, but have become the end. Their passionate practitioners believe it is impossible to do proper economics without diagrams; that diagrams are not only necessary but also sufficient. You cannot say you have solved the problem until you have drawn the diagram. And when you have drawn the diagram, you have done all that needs to be done. The aim seems to be to express all economic problems as problems in Euclidean geometry, and all capable of solution. It is believed that if the problem cannot be expressed diagrammatically, it cannot be solved; and once it is expressed diagrammatically, it is solved. To diagrammatize a problem, is to solve it. Every new diagram is seen as another step forward in the development of economics; the diagram is a sign and symbol of progress, and there can be no progress without diagrams.

Although the diagrams in this book have been neatly drawn and nicely described and some have been given an account of their genesis, there seems not to have been a serious attempt to subject them to a *cui bono* test. It is certainly true that **indifference curves**, for example, are "a fundamental expository tool of economics", if by economics we mean "economics as taught in modern textbooks," but would economics have suffered if indifference curves had never been invented? To push the utilitarian criterion even further, have economic conditions generally benefited from their invention?

Critics will ask whether diagrams are a help or a hindrance to the progress of economics, but for convinced diagrammatizers this is a non-question. In their view, economics is diagrams; the diagrams contained in this publication constitute the intellectual structure of economics; the ability to comprehend these diagrams is what differentiates the economist from the non-economist.

Geometry is fun, for some; and those who think otherwise should tolerate and respect it as a pleasant pastime, but what the anti-diagram people do not like is a tendency of diagrammatizers to command the high intellectual ground; to strive for hegemonic domination of the curriculum; to claim that modern economics is essentially mathematical and diagrammatical; and that any other approach is sub-scientific and not worthy of being called proper economics.

There is no doubt that some students enjoy diagrams. They encounter diagrams at their first contact with economics, and come to expect them at every stage of their economics education. If they have done well in reproducing diagrams in essays and exams, they are disappointed if lectures are devoid of them. They go away feeling disillusioned and cheated. Diagrams also have a therapeutic function. They convey a sense of Euclidean certainty, of academic respectability and of scientific precision, thus providing a security blanket in an uncertain world, and satisfying a psychological need, even though economics itself has little or no need of them. But does a liking for diagrams prove they are a necessary part of an economics education, or is it merely recognition of the power of geometry to provide intellectual satisfaction? Are we teaching students economics or geometry? For as long as geometers continue to exist, some will not be able to resist the temptation to turn economics into geometry, confident they are making it more scientific, and ridding it of the vague and ill-defined tools of communication known as words.

The attraction of diagrams can be attributed, at least in part, to the fact that with television, and telephones that provide pictures as well as words, and advances in graphic arts, we live increasingly within a visual culture, which in itself is perhaps no cause for complaint, but which becomes worrying when it begins to erode and replace verbal culture. Students in lecture rooms will sit, more or less patiently, more or less politely, waiting for the words to stop, with pen in hand, without taking a single note; but as soon as a diagram appears on the screen, the pens come into action. Words are regarded as unimportant preliminaries to the important diagrams.

The editors of this collection are of the view that multi-coloured diagrams are being used "to great effect" in modern textbooks. Perhaps Marshall's warning that diagrams are "merely pictures" would have been less dismissive if he had seen them in glorious technicolour. Those who dislike the use of diagrams in economics would regard their coloured proliferation as yet further evidence that Picture Book Economics has become a substitute for verbal analysis; and that in economics words are being submerged by pictures. Textbook writers compete with one another to see who can include the most diagrams in the most colours.

Part II: Useless and redundant diagrams

Turning attention now to an assessment of some particular diagrams in this collection, it could be argued that, when subjected to critical appraisal, some are theoretically deficient; some seem to be useful in theory, but useless in reality; others are redundant, either because they are too trite to be taken as serious contributions or because the concept being portrayed has been adequately enunciated long before the diagram was invented.

A rare contribution in this collection to have been given a non-eulogistic treatment, and to have received an objective appraisal of its validity and usefulness is that of Yew-Kwang. His six qualifications to the **Harberger Triangle** raise serious doubts about its academic status.

They amount to a denunciation of the Harberger Triangle as a valid tool of economic policy. Although Yew-Kwang does not state that conclusion in his six qualifications, he argues that even if taxation means a loss of surplus "it does not follow that there should be no taxation or less taxation", because the tax revenue could finance public goods which create consumer surpluses and offset the loss caused by taxation.

A similar tone of dissension is evident in the contribution by Richard Lipsey, who criticizes the Harberger diagram for "paying no attention to the effect of policy changes on the distribution of income", and who believes that many economists "prefer conclusive results that suppress inconvenient truths over inconclusive results that take account of these truths".

A critical, as well as expository, approach was also taken in the contribution on **cobweb diagrams**. The author's incisive assessment concludes that the cobweb diagram is "fatally flawed as a theory of agricultural price movements", and is not supported by empirical evidence. It has nevertheless had a "profound and lasting influence" on the development of economic models. This must surely be convincing proof of the power exerted by diagrams, and their founders and teachers, in influencing curriculum design; a power that persists irrespective of the intellectual and empirical merit of the ideas the diagram is purporting to illustrate.

The enthusiasm that surrounds some diagrams and the exalted status they enjoy in the corpus of economic theory seem to be out of proportion to their role as progenitors of economic wellbeing. These diagrams could be categorized as useful in theory, useless in practice. The **homothetic function** is one such. It is said to be "important to international trade", but surely this confuses trade with the theory of trade. But it could also be said that international trade itself would not suffer if homothetic functions did not exist; and that when businessman are negotiating an import or export deal, or when government officials are making decisions that affect exchange rates and the balance of payments, it is very unlikely that any of the decisions would be based on homothetic functions.

Offer curves could also be placed in this category. They are said to be "immensely useful" to international trade theory, but how useful are they to trade? Does anyone actually engaged in international trade rely on them? One suspects that their main function and reason for being is as a starting point for the development of even more sophisticated diagrams of trade, not for the development of trade itself.

The same could be said of **cost curves**, described as "a staple part of the curriculum of undergraduate microeconomics". Expertise in drawing and manipulating cost curves will be well rewarded in academic examinations and academic careers, but how many company directors draw cost curves before making price and production decisions? Once they escape from there neoclassical education, they discard the toys of their youth. If diagrams are so important in economics, how do you explain that with the exception of school and university teachers, they are rarely used in the real world? Rather than enhancing economics, it is possible that the diagrammatic approach, so beloved and vaunted by some, has in fact contributed to current disillusionments with economics.

The **optimal tariff diagram** provides yet another example of the futility of much diagrammatic endeavour in economics. That diagram is said to appear in "nearly every modern textbook on the theory of international trade", and to provide governments with "the necessary formulae" for setting an optimal tariff to maximize their economic well-being. The ingenuity and

cleverness of the diagram can only be matched by its uselessness, in a world moving towards free trade and the abolition of all tariffs. The futility of the diagram is even more pronounced when it is acknowledged that, if two trading partners adopt an optimal tariff, they could both possibly be worse off than under free trade. This diagram persists in modern textbooks because of its useful role as a mental exercise and student minding activity.

There are diagrams in this collection which could be labeled superfluous or redundant. One reason for their superfluity is that the words explaining the concept behind the diagram are just as clear as the diagram, or even clearer. Any attempt to praise or defend the role of diagrams as expository devices in economics would need to show that diagrammatic presentation is superior in that respect to verbal explanation. If it cannot be shown to be even slightly superior, does this not mean that the diagram is a waste of time, effort and space? If the verbal explanation is not expressed clearly, the diagram could well be superior as an expository device; but if the verbal explanation can stand alone as an adequate means of communication, the diagram becomes a mere visual embellishment, lacking any substantive function. A case in point in this collection is the chapter on **backward-bending labour supply curves** by John King. This knowledgeable account of the history and significance of the concept is so lucid, it renders its accompanying diagram redundant.

Another reason for placing the backward-bending supply curve of labour in the redundant category is that the concept underlying it is far from new. The concept was not brought into being by the curve. It was, for example, stated as long ago as 1798 by Malthus, who announced it not as a triumphant discovery, but as a commonsense observation. In discussing proposals to increase the level of poor relief, he said: "the receipt of five shillings a day, instead of 18 pence, would make every man fancy himself comparatively rich, and able to indulge himself in many hours or days of leisure. This would give a strong an immediate check to productive industry" (*Essay on Population*, 1798). Has the portrayal of the backward-bending supply curve of labour contributed any advance to the progress of economics, or did Malthus in 1798 say all that needs to be said on the matter?

Malthus might also be cited in reference to the **Kuznets Curve** and the **Laffer Curve**. These two curves add little to the concepts they purport to explain, and these two concepts a little more applications of the methodological principle called by Malthus the "doctrine of proportions", according to which the relationship between a determining variable and the determined variable in economics is, more often than not, parabolic, and the optimum position is a balance or middle way between extremes.

The **UV curve**, showing figures for unemployment and vacancies, is another example of a curve that is quite redundant. It is obviously important to know whether unemployment figures are higher or lower than vacancy figures, with the former situation indicating a slower demand for labour and an actual or potential recession, and the latter situation, indicating a high demand for labour and an actual or potential expansion; but once these statistics are known, does the drawing of the curve add anything of theoretical or policy significance? Its function would appear to be merely decorative.

The **rent seeking diagram** is yet another instance of diagram redundancy. The entry in this collection provides an exposition of what is meant by rent seeking and of how government attempts to regulate the free market might involve a high expenditure of resources without a proportionate increase in output. The verbal argument may or may not be convincing, but its plausibility does not appear to be in any way enhanced by the accompanying diagrammatic

argument. The diagram seems to be designed to attach academic respectability to a thesis, in a world where academic respectability is measured by the use and complexity of diagrams.

The same could be said of the **circular flow diagram**. As Roger Backhouse points out, the concept of circular flow lends itself admirably to a diagrammatic exposition, and this has been encouraged recent years by the development of a "visual culture" throughout the wider community, and by improvements in the art of graphic design. The textbook ubiquity of the circular flow diagram is evidence of its popularity and usefulness as a teaching device, or but this is probably as much due to its simplicity as to its intellectual content. Some would say that, by comparison with most of the diagrams in this book, it is a chart rather than a true diagram. The concept of circular flow is simple, obvious and easy to explain. Its chart is surplus to requirements.

Diagrams can also be deemed superfluous because of the trivial nature of the concept or theory being diagrammatized. Some of the diagrams in this collection and some of their underlying concepts seem to be less significant than claimed or suggested. Some indeed are so obvious as to be best described as trite. **Engel's Law** might be one such. Do we need a law and a diagram to know that expenditure on food declines as a proportion as household total expenditure increases, or that family size and composition affect expenditure patterns. If Engel's Law had not given rise to a diagram, it would probably never have been deemed worthy of mention. Its fame as a proposition and its position in the curriculum and textbooks of economics seem to have come from its diagram.

As a final example of an unnecessary diagram, we could refer the **IS-LM diagram**, which after generating an industry of academic commentaries now seems to have been rendered redundant, given that its creator acknowledged its limitations and expressed amazement at the amount of attention it has generated. It is a sobering reminder of the power of diagrams to live flourishing and self-perpetuating lives of their own, independently of the intellectual merit or validity of their underlying concepts.

Part III: An anti-diagrammatic conclusion

The modern economist's preference for diagrams over words would have been understood and applauded by the professors of the School of Languages in the Grand Academy of Lagado in Balnibarbi. In Jonathan Swift's *Gulliver's Travels* (1726) the professors aim to improve their language by abolishing words and by using implements to express themselves. Each person would carry sacks of implements, and converse with one another by holding up the necessary implement. As Gulliver noted, the inconvenience of this system is that you could be obliged to carry great sacks of things on your back, unless you could afford strong servants to carry them. The professors of Balnibarbi had apparently not thought of using diagrams as substitutes for words.

Does this collection of diagrams represent the benchmark of competence in economics, and the standards that our students should aim to attain, so that they too can be deemed worthy of the name "economists"? Or is it sad evidence of how economics has lost its way and become swallowed up by geometry? Do these diagrams constitute the high point in the evolution of economics, or are they a pointless exercise, a party game, a cosmetic embellishment, or the decorative bunting of economics?

But for all enthusiasts of diagrams and for all who see their role as inspiring the next generation of economists with a love of diagrams, so that they in turn are inspired to create even more diagrams and to shunt the car of economics even further on to the diagrammatic track, this publication will be a memorable iconoplastic achievement.

Author contact: jpullen2@une.edu.au

_____

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27rwer-issue-65/

A plea for reorienting philosophical attention from models to applied economics

Gustavo Marqués   [University of Buenos Aires, Argentina]

Criticism of traditional economic theory

In the past, conventional economic models have been criticized for their inability to explain and predict, as well as for the difficulties in applying them to specific economies. It was assumed that these shortcomings were due to the "unrealistic" nature of some of their assumptions, and for decades this reason has been advanced to dismiss their practical utility and the possibility that they could be true. Musgrave (1981) makes an important contribution to this debate showing that many of the statements used in the formulation of economic models – which were considered false when interpreted in a too literal sense – properly understood mean something completely different that may also be true. A similar approach is advocated by Lipsey & Steiner, 1981.[1] Recently, new arguments have been offered that improve our understanding of the role of different classes of stylized assumptions (i.e., idealizations, distortions and omissions) in the construction of economic models (Mäki, 2002, 2008, Weisberg, 2007). These contributions have been successful in addressing the old-fashioned "realistic" attack on conventional economic models, showing that they cannot be dismissed as carriers of relevant information about the world on the basis of these reasons, and there is nothing inherently wrong in the practice of *mainstream* model building[2].

However, these arguments relate only to the question of the truth-status of economic models, not to their practical utility. Their nature is, so to speak, *negative*: they are arguments designed to criticize the arguments of the critics. Although they are able to defeat the rather naïve objections of the old-fashioned critics, they fail to dispel the concern that many practicing economists feel regarding the questionable performance of economic models. Since the beginning of the XVII century, many economists of different orientations have expressed their dissatisfaction with the course adopted by economic theory and their concern about its regular procedures. It is not a coincidence that Mill's influential defense of the scientific character of economics focused on its "abstract" nature, taking as granted that it could not be successful in the domain of real world ("concrete") phenomena.

Apparently, this discomfort regarding the merits of economic theory is still felt, despite the efforts of current *mainstream philosophy of economics* (MFE), which attempts to exhibit the rationality of mainstream economics' practices and provide epistemic legitimacy to standard

---

[1] "Consider a theory that assumes the government has a balanced budget. This may mean that the theorist intends that theory to apply only when there is a balanced budget; It may *not* mean that the size of the government's budget surplus or deficit is irrelevant to the theory"(Lipsey and Steiner, 1981, p.24).

[2] Maki's rejection of traditional criticisms of economics asserts that economic models, even those that isolate ideal mechanisms, *can* be true. This is a *possibility* because in fact he makes no claim that these models are indeed true. Maki's arguments are presented in the form of "even-if arguments", stating that, even if certain conditions in the formulation of a model are met, leading us to say that such models are false, they may still be true (see Mäki, 2008). Statements of this kind are the argumentative core of Mäki's minimal realism. They offer a critique of the most common objections to the traditional way of modeling, and in this sense they are just *negative* arguments in the sense intended here. To the extent that these arguments are successful, they only show that the reasons usually directed against the epistemic relevance of traditional economic models are not good for that purpose.

models. See, for example, the words used by Mäki at the beginning of his *Facts and fictions in economic theory*:

> Fact or fiction? Is economics a respectable and useful reality-oriented discipline or just an intellectual game that economists play in their sandbox filled with imaginary toy models? Opinions diverge radically on this issue, which is quite embarrassing from both the scientific and the political point of view... Economics is a contested scientific discipline. Not only are its various theories and models and methods contested but, remarkably, what is contested is its *status as a science*.... Suppose we take one of the characteristics of science to be the capability of delivering relevant and reliable information about the world. Suppose furthermore that this is not just a capability, but also a major goal and actual achievement of whatever deserves to be called by the name of 'science'. How does economics do in this respect? This question is about as old as economics itself (Maki, 2002, p. 3).

His stance is rather odd, because these questions do not arise in the case of other sciences. Who might seriously doubt that physics or biology do provide "relevant and reliable information about the world"? Therefore, it seems that there is a particular difficulty for making sense of theoretical practice in economics. In particular it has not yet been properly clarified in what sense it is applicable to real economies, although this has been the focus of much debate for more than two centuries. Apparently, *negative* philosophical reasons suffice to get rid of the usual criticisms of mainstream economic models. But these reasons are not sufficient to help economists feel proud of the concrete results of their models, and for philosophers of economics to provide a convincing justification of the capability of conventional economic models to suggest and underpin economic policies. To achieve these goals additional (this time *positive*) reasons should be offered.

Let me be clear about the nature of the positive reasons that are needed. They belong to two main kinds:

> a) Some of them could be advanced in order to showcase where a particular content of economic models provides understanding of some aspects of real economies. Apart from exceptional cases of very appealing models (like those of Shelling and Akerlof that have been frequently commented on in recent literature)[3] this issue remains largely unclear once the bulk of economics models are considered. Even in the case of those authors that take for granted that economic models offer valuable information about real economies, the nature of this information remains poorly specified and the merits of those models largely unargued. Regarding concrete economies it is still unclear what exactly is learned from models, and why that which is supposedly known through them cannot be learned by other means (venerable traditional theories, common sense or *practical* economic knowledge).

---

[3] See Akerlof, George A. (1970). The market for 'lemons': quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84: 488-500, and Schelling, Thomas C. (1978). *Micromotives and Macrobehavior*. New York: Norton.

b) More importantly, given the problems that assail the empirical testing of economic models and the general problem of external validity, what is mainly required is to show concrete cases of successful economic technologies able to generate unequivocal practical results (I mean, results that cannot be reasonably attributed to other causes than the insights provided by models). If as MFE asserts economic models are the kernel of economic wisdom and the engine of economic progress, philosophical attention must be applied to show where (in which particular cases) and how they contribute to the acquisition of practical results. Physics, biology, genetics and some other undisputed (and reputed) scientific practices are mainly preached not by their successful tests, but by their capability to mold and transform our daily life through related technologies. Can the practice of economics' model building be defended on these grounds?

The insufficiency of positive philosophical arguments to sustain the representational as well as the practical usefulness of mainstream economic models is to a large extent the result of a feature of these models, which relies on a substantial use of a particular type of assumption, called *tractability assumptions* (Cartwright, 1999; Alexandrova, 2006).

This paper examines some of the major solutions that have been proposed in order to avoid the trade-off between the use of tractability assumptions and the external validity of the results that are obtained with their help. It is argued that these contributions have failed both in clarifying what is the usable real world-oriented content that economic models offer, and in showing clear instances of successful applications of economic models (i.e., economic technologies). It is also claimed that these new sympathetic approaches to highly idealized economic models fail to exhibit what exactly are their particular contributions in those applications. Worse, it has been claimed that those solutions rely on a different approach, which enhances the crucial role of background (extra-modelic) knowledge, something that seems to be difficult to accommodate within their shared view of economics as a model-centered discipline. What is needed, we suggest, is to shift philosophical attention from the conundrums of representations to the conditions that must be fulfilled for building a substantial core of successful applied economics.

The problem of "overconstraint"

In "The Vanity of Rigour in economics", Nancy Cartwright examines a special type of economic model, called by Lucas *Analogue Economies.* There is no doubt that these "economies" are *unrealistic* in their construction, as Lucas himself explicitly recognizes: "Any model that is well enough articulated to give clear answers to the questions we put to it will necessarily be artificial, abstract, patently 'unreal'"(Lucas, 1981, p. 271)".

How damaging is this circumstance for the aspiration of applying the results of a model to situations of our world? In the past, Cartwright defended the unrealistic nature of the models within the framework of a taxonomy that recognizes two types of models: those which "establish facts about what is happening in the real economy" and those that "establish facts about stable tendencies". She has argued that the *analogue economies* belong to the second type.[4] We will refer to them as tendency-models.

---

[4] "....we do not need to assume that the aim of the kind of theorizing under discussion is to establish results in our analogue economies that will hold outside them when literally interpreted. Frequently what

Tendency-models are designed to isolate the action of a single cause and show what its "pure" contribution to the generation of a brute event, which "happens" in real economies, is. In Cartwright's terms, tendency-models do not describe facts at all; they describe the *exercise* of a capacity, not the *result* of this exercise. Now, if the goal is to capture a tendency, "it is essential that models make highly unrealistic assumptions, for we need to see what happens in the very unusual case where only the single factor of interest affects the outcome" (Cartwright, 2007b, p. 219).This result is important because it leads to the conclusion that the reproach of "unrealism", which traditionally has been used to explain the lack of successful applications suffered by conventional economics, is in these cases misplaced.

However, Cartwright's thesis that through isolation it is possible to identify capacities is more a response to the traditional way of objecting to the practice of modeling in economics than a defense of conventional models; because according to Cartwright *it is* possible to direct against *analogue economies* a more sophisticated criticism, showing that *there is* a problem after all in this way of modeling. Although all models distort reality, this effect may be due to two basically different strategies: to *omit* some factors present in the target system or to *add* to the model factors that are absent in the target. Subtracting and adding are very different activities.[5] The use of additions can endanger Cartwright's usual strategy for defending the practical relevance of "unrealistic" economic models, which asserts that they reveal the "pure" contribution of a cause (or a set of causes) to an observable effect in our world. To see where the difficulty lies, it is useful to distinguish two types of "idealizations" in economic theory:

(a) Galilean idealizations, which omit any possible interference to the action
of an isolated cause;
(b) Non-Galilean idealizations, which are used to incorporate within the model
features that do not have any counterpart in the targeted real economies.

The first are beyond reproach, according to Cartwright, since they are necessary to find out "capacities".

Let us call this kind of idealization that eliminates all other possible causes to learn the effect of one operating on its own, Galilean idealization. My point is that the equivalent of Galilean idealization in a model is a good thing. It is just what allows us to carry the results we find in the experiment to situations outside – in the tendency sense. "We need the idealizing assumptions to be able to do this". (Cartwright, 1999, p. 12).

The idealizations of the second type are necessary to reach (deductively) precise and well-defined results. They are, however, a source of problems because they exacerbate the trade-off between internal and external validity. Cartwright offers two arguments to justify why Non-Galilean assumptions are problematic in this regard. The first one focuses on the *amount* of this type of supposition. She holds that economic models "are complex, at least by comparison with physics models doing the same kind of thing: they have a lot of structure. The list of assumptions specifying exactly what the analogue economy is like is very long"

---

we are doing in this kind of economic theory is not trying to establish facts about what happens in the real economy but rather, following John Stuart Mill, facts about *stable tendencies*" (Cartwright, 1999, her italics).

[5] It seems counter-intuitive, but a simplified world simpler than the actual one can be represented within a model either by omitting or adding factors. Additions can cause special problems, however. This is particularly true for models of trends, which assume that the objective of the model is to exhibit the "pure" or "natural" capacity of some factor.

(Cartwright, 1999, p. 15). However, the main problem brought out by the incorporation of Non-Galilean assumptions comes from their non-representational nature, which is conditioned by the goal that these assumptions are supposedly helping to reach: precise results by perfectly deductive means. Galilean Idealizations respond to the interest of providing a simplified representation of reality; Non-Galileans Idealizations, on the other hand, introduce rather arbitrary specifications just for the purpose of allowing or facilitating inferences and achieving consequences with deductive accuracy. In the first case factors which are supposed to be present in reality are omitted, while in the second factors which are regularly assumed not to have a counterpart in reality are nonetheless included.

What is then proved in this kind of model is that factors C isolated within them generate a result R in the presence of (exceptional) conditions N, which are posited with the only purpose of reaching R with deductive rigor. But then the contribution of C to the generation of R in the framework of the model does not give us information about what would be the contribution of C to the generation of R in real (concrete) economies, in which N supposedly are not present. It cannot be now assumed that the *capacity* of C, discovered within the model, is an ability that C would still have whenever it is operating out of the model (in real economies):

> What I want to talk about today is a problem that can beset real and thought experiments alike and in both physics and in economics. But it is a particular plague for thought experiments in economics, I shall argue, so much so that it regularly undermines the use of models to establish capacity claims. That is the problem of *overconstraint* (Cartwright, 2007a, p. 73).

The problem with tractability assumptions

As we have just seen Cartwright considered that the phenomenon of "overconstraint" puts into question the applicability of economic models to situations of our world. But it may be thought that this difficulty has only a limited impact and it occurs only in the framework of the ontological assumptions made by Cartwright, in which models are designs aimed at discovering tendencies or capacities. As long as this ontological commitment is controversial, it is important to describe the nature of the difficulty in a more general way, outlining its logical dimension.

Suppose that B describes a relevant and desirable result and A describes a circumstance whose presence in our world is attainable. At the moment, there is no known logical connection between A and B. Now suppose that someone asks what set of additional assumptions would allow one to deduce B from A (just using the ordinary rules of logical construction and derivation). An ingenious individual puts his mind to work and finds that assuming an arbitrary set of assumptions C (which are only restricted by logical considerations) A implies B. Here ends, successfully, the exercise. Conditions C are mere assumptions in the logical sense of the term: they are starting points for the argument. Following Kuorikoski and Lehtinen, (2009) and Kuorikoski et al, (2010), we call them *tractability assumptions*.

What does this demonstration prove in reference to *our* world? More precisely, what is the relevance, if any, of such an argument regarding the applicability of the results thus obtained to real economies? If prior to its construction, we believe that no causal connection between A and B exists in our world, why would the demonstration provided in this exercise contribute to

changing our minds? A reasonable response could be that C describes conditions that are *plausible in our world*. But by hypothesis, it is assumed that this is not the case. As a result, the derivation does not contribute to the credibility that a causal link between A and B exists. It seems that we are facing a dilemma. If previous to the logical exercise we do believe in the existence of a causal nexus between the aforementioned factors its rigorous demonstration does not add anything to our conviction; but if at the beginning we do not believe in the existence of a causal connection between A and B the exercise does not compel us to change our mind. What then is gained by the proof of the connection between A and B given that it has been obtained at the expense of extraordinary (can I say "unrealistic"?) circumstances?

The presence of tractability assumptions poses a problem for those who wish to defend the epistemic relevance of economic models. Two types of solutions have been proposed. The first, which we call "internalist", argues that certain operations carried out within models (i.e., inside what Sugden calls the "model world"), particularly derivational robust analysis, may show their epistemic credence[6]. We find this attempt unsuccessful and misleading, but in this paper we will not examine this claim. "Externalist" solutions, on the other hand, argue that in order to acquire epistemic relevance economic models have to be supplemented with some kind of external knowledge. In the following sections three different strategies to sustain this view will be examined: (a) the interpretation of models as parables (Cartwright); (b) the suggestion that what is needed is to train suitable interpreters (Colander), and (c) the concept of models as open formulas (Alexandrova). Despite their differences, all of these views are based on the assumption that economic models including arbitrary tractability assumptions contain reliable and relevant knowledge about our world. The problem, according to these views, is that it is not directly usable: the applicability of economic models to situations of our world crucially depends on the assistance of some kind of background knowledge coming from outside the models themselves.

Cartwright's vision of models as parables

As seen above, Cartwright (1999, 2007a, 2007b) called attention to the problem of the "overconstraint" suffered by economic models of the "analogue economies" type, which generated a trade-off between their internal and external validity. In a more recent paper (Cartwright, 2008) she offers a new interpretation and argues that the trade-off may not take place after all. To reach this conclusion she contrasts two ways of understanding economic models: as fables and as parables. She argues that fables deliver a "message" or "lesson" that is explicitly formulated within the text. Parables, however, shed (or perhaps it would be better to say "suggest") a lesson, that is not contained in the model itself, but must somehow be built from the outside taking into account relevant portions of available background knowledge. This means that models can have a "correct" lesson within them, but it must be partly construed out of the materials provided by the model on the basis of theoretical and extra theoretical knowledge. Models deliver a lesson that despite being abstract in nature may

---

[6] There are two main versions of this position. On the one hand the autonomist view of Knuuttila who reconsiders the concept of epistemic relevance, untying it from any reference to our world. On the other hand, derivational robustness analysis, as understood by Kuorikoski, Lehtinen and Marchioni (2010), replaces the comparison between a model and its intended target by the comparison between different versions of a basically identical model. In this case the strategy is to build a family of models and claim that the derivational robustness analysis allows to identify existing causal connections in our world simply by comparing the members of the family.

be applicable to the specific economies of our world. Her vision of models as parables can be understood as a new strategy in the broad project of mainstream philosophy of economics intended to "recover the practice" of mainstream model building.

Cartwright's new vision is consistent with an idea that she has advanced before (see for instance Cartwright, 2007b): economic models do provide valuable informative content, and if there is any doubt about what their epistemic relevance is, the problem lies not in the models themselves, but in our difficulties for developing *another* type of knowledge able to reveal to us *how to use* a models' information. This is precisely what happens with parables. To identify their "lessons" and be able to apply them to situations of our world the use of background knowledge is crucially required. Her defense of the epistemic relevance of economic models whose results depend on the discretionary addition of tractability assumptions is unconvincing, however. Let me mention some problems of this vision.

(1) There is no guarantee that such models will deliver a "correct" abstract lesson (i.e., a lesson applicable to the real world). On the other hand, even if models contain materials for building the right lessons out of them, there are no rules for identify them unequivocally. Besides, the lessons that models could suggest, being dependent on the particular state of knowledge which prevails at the moment, may vary according to times and places, and are always subject to revision.

(2) The lessons and applications that models facilitate are no longer based on the *consequences* obtained in the model but on other, more abstract content, which is not deduced from the model, but is "inferred" or "captured" otherwise. Arguably, then, the problem of overconstraint is not resolved, but is rather diluted by changing the reference point: the *consequences* are still over-constrained (since Cartwright is not advocating a change in models, but in their interpretation), but now the focus is placed not on them (or their applicability), but in *lessons* which supposedly do not depend on the set of tractability assumptions.

(3) To spread their message, parables do not need to deliver rigorous proofs, and much less have recourse to the employment of advanced math or heroic idealizations. In fact, it seems that the "lessons" that economic models offer could be obtained without having to impose deductive power within the model by adding strategic tractability assumptions. Why do modelers send messages or lessons through analogue economies? If the epistemic value of models resides at a more abstract level, what is the purpose of over constraining their results (often with many tractability assumptions)? It seems that a more informal argumentation would be enough (and surely the lesson so delivered would be clearer).

 (4) A potential problem of Cartwright's shift from methodology and epistemology to literary analysis is that parables, as many everyday sayings, are not only ambiguous in their content, but frequently suggest opposite lessons that contradict each other. One can then choose the lesson that best suits his interests or the particular occasion. This pliability of the parables could certainly explain the ease with which applications for economic models are found and their epistemic relevance taken for granted. I can't tell whether this plasticity should count as a credit or a defect.

Interpreting economic models

In his article "The economics profession, the financial crisis, and method", David Colander focused his analysis on the performance of "the dominant dynamic stochastic general equilibrium macroeconomics model" (DSGE) regarding the global financial crisis of 2007-2008. This crisis was so deep that "the world economy came perilously close to a systemic failure in which a financial system collapse almost undermined the entire world economy as we know it" (Colander, 2010, p. 1). In this case his analysis refers not to "analogue economies", but to a type of model designed to be applied to a particular situation of our economic system. It is then interesting to see how its performance is evaluated. Colander holds a moderate position, pointing out, like Cartwright did regarding analogous economies, that such a model has valuable information about the world which deserves to be considered and elaborated[7].

As usual only shortcomings are explicitly mentioned. In this regard he argues that those who were looking at the world through DSGE's lenses were prevented from seeing that conditions for the advent of the crisis were growing inside the economy, despite that "the possibility that a crisis might occur at some point was becoming evident to many observers"[6]. To some degree one could excuse this failure pointing out that after all predictions are usually unattainable in economics. But Colander emphasizes a rather different point. He asserts that "it did not take a rocket economist to recognize problems in the financial sector as the burgeoning sub-prime mortgage market was bringing in less creditworthy buyers. At some point that process of credit expansion had to end". This observation seems to imply that those who did *not* have the help of DSGE's analytical tools had a *clearer* perception of the situation than those who counted on the model's help. Leaving aside the difficulties of anticipating future events, the fact remains that this model has also been useless to *analyze and understand* the crisis *once it was already present*.

The inadequacy of the model for examining the crisis is explained by the purpose of simplifying its object of analysis, which exhibits a substantial complexity. To be tractable "the DSGE model ruled out meaningful considerations of the financial crises by its representative agent and global rationality assumptions". Colander does not have much hope in the strict adherence to this strategy. In his opinion "mathematical modelers should deal with that complexity head on, rather than focus on models that assume much of that complexity away as we believed the dominant dynamic stochastic general equilibrium (DSGE) macroeconomics model did" (Colander, 2010, p.1). Consequently, Colander rejects the ongoing practice of working on models like the DSGE and advocates for developing more sophisticated models, which are characterized as "highly complex heterogeneous agent, coordination failure models that might have been able to incorporate such events as a crisis of confidence".

It could be thought that this (forthcoming) new generation of highly complex models would finally meet the requirement of epistemic relevance. However, Colander admits that future models *will not provide a firm basis for the implementation of successful economic policies* either. As he points out, "models of complex systems do not, and at our current stage of knowledge, cannot, provide definite policy answers – they simply provide guidance to individuals who have real-world experience and a detailed knowledge of the institutional structure".

---

[7] As he says, the criticism that DGS has received "does not mean that such abstract modeling should not be done; We strongly supported such basic research" (Colander, 2010, p.421).

Colander's view is convergent with the ones offered by other authors that we examine in this paper, like Nancy Cartwright and Ana Alexandrova, in the sense that they all believe that standard economic models do provide some type of useful and enlightening knowledge, even if this knowledge is not directly linked to situations of our world. Therefore, in all these cases the relevant question is *how that knowledge should be used to obtain practical results and build successful economic policies.* What distinguishes Colander's views from other opinions is the nature of the proposed solutions, which consist in this case in academic and attitudinal changes.

First, it is advised that economists assume more responsibility when the properties and results of the DSGE model are communicated outside the narrow community of experts and model builders. Particularly, published models should include "an explicit warning directed at the non-scientific users of the model. This warning could include a list of what the researchers see as limiting assumptions of the model, and the researchers' beliefs about whether the model can be used to guide policy" (Colander, 2010, p. 424)

Still more important is his indication that expertise in the use of macroeconomic models requires practical knowledge of the economy as well as other kinds of knowledge and skills, which are different from that involved in the practice of modeling. In particular, Colander proposes a crucial institutional shift, which consists in allocating public funds for training economists in the interpretation of models with a view to their applications.

> Currently, most economists are not selected for their ability to, or trained in how to, 'choose' an appropriate model, or otherwise relate a model to policy. Doing this requires knowledge of a wide range of models, historical knowledge, and institutional knowledge. They are trained almost entirely to produce models. The other ability they must learn on their own. By design graduate training has eliminated those courses, such as the history of economic thought, methodology, economic history, or courses surveying literature, that are most relevant for training students to choose among, and interpret models...A potential solution to this problem is to increase the number of researchers trained in interpreting models rather than developing models. This would mean viewing the interpretation of models as a separate skill from producing models. If a funding agency were to provide research grants specifically to interpret models, that problem could be somewhat alleviated. In a sense, what I am suggesting is the creation of an applied science subdivision of the National Science Foundation's social science division. This subdivision would fund research on the usefulness and interpretation of models. (Colander, 2010, pp. 425 – 426)

Colander's proposal is original and interesting, but somewhat understates and dilutes the role which, according to mainstream philosophy of economics, models play in the production of relevant and reliable knowledge about the specific economies. Colander's perspective makes the potential usefulness of this knowledge heavily dependent on the acquisition of other kinds of knowledge whose source and legitimacy was originally contested. In fact, Mill, Senior and many other economists of the past claimed that economic theory was scientific in the sense that it went well beyond the knowledge of economic affairs available to ordinary people, historians of economics and entrepreneurs. But it seems that Colander's perspective re-

enhances the role of the very kinds of knowledge that were thought superseded by economics' theoretical practice.

Technological use of economic models. Towards a more applied economics.

Recently, Ana Alexandrova has defended a new vision concerning the role that economic models play in the implementation of economic policy. In it she limits her analysis to one particular model that is praised as a paradigmatic case of successful application in the design of institutions, the auction model (Alexandrova, 2008; Alexandrova and Northcott, 2008). Her purpose is to explain what its contribution is in the achievement of this goal. She argues that the main existing rival views about models' applicability are not useful in this case. Alexandrova's approach is a promising way to defend the practical relevance of economic models, suggesting, at the same time, a more general way to appreciate what exactly the applicability of models that incorporate tractability assumptions depend on.

Economic models can be used to represent (and be applied to) a certain target T. According to Alexandrova, there are two main views that seek to give an account of their applicability: the "satisfaction of assumption" account, which is attributed to Daniel Hausman, and the "capacity account" developed by Nancy Cartwright. According to them, a model represents (and is applicable to) T when, respectively, its assumptions are satisfied in T or the causes described in it occur in T. To illustrate her position, let's express it in Hausman's concept: If a model M contains assumptions, some of which are idealizations (we read: tractability assumptions), this fact prevents them from being strictly true in T. But in that case, according to Hausman, it is possible to gradually de-idealize those assumptions until they match the relevant characteristics of the "intended target". De-idealization allows models to be applicable and to acquire empirical content.

Alexandrova (2006) points out that this strategy is only possible in some cases because it is not always possible to de-idealize the tractability assumptions incorporated into a model. She says, for instance:

> In what sense is it more realistic for agents to have discretely as opposed to continuously distributed valuations? It is controversial enough to say that people form their beliefs about the value of the painting or the profit potential of an oil well by drawing a variable from a probability distribution. So the further question about whether this distribution is continuous or not is not a question that seems to make sense when asked about human bidders and their beliefs (2006, p. 183).

Her proposal is then intended to "recover" the practice of model building in those cases in which de-idealization cannot be followed. She tries to give an account of how, despite this limitation, economic models can be applied successfully to obtain desired economic institutions and practical results. To examine her vision let me describe a model M in this sketchy way:

> Given C1,..., Cn, a certain characteristic F causes behavior B   (1)

A more synthetic way of expressing its content is:

"Under conditions C, F causes B" (1')

Note that M is a *closed* formula, in the sense that all of its assumptions are specified. But one thing is the model and another its use. Alexandrova points out that M can be *used to build a hypothesis* in which only *some* of its assumptions (or *none* of them) are specified. The hypothesis has this form:

Under conditions X (that *may or may not* include conditions C), F causes B (2)

In (2), C has a purely notional presence, since it may be completely undetermined. For this reason she proposes considering models as *open formulas*. Strictly speaking, the content of (2) boils down to the following:

(In our world) There are conditions X, where F causes B (2')

In (2') F and B are conditions whose properties are known and X is the unknown variable whose "values" have to be found. Interestingly, the original model, which suggests the hypothesis (2) and (2'), does not provide any clues for discovering those values. In fact, the model itself contains no hypotheses such as those made in (2) or (2'). They are independent from the model, though inspired by it. From this point of view the model has no real world informative content of its own: it is rather considered as a source of hints, tools and resources for generating hypotheses about the world. In Alexandrova's words an auction model functions as a "framework or heuristic for formulating hypotheses".

But how workable is such a heuristic? Is it really a form of heuristic after all? Suppose that "B" is a desirable outcome and "F" is a state of things, which we can implement in reality. Suppose then that a model M proves that under conditions C, F causes B. C describes a set of conditions that are logically *sufficient* to ensure such an outcome. The epistemic significance (relevance) of the model seems to depend strongly on the feasibility of conditions C. What is then accomplished by rigorously proving that "F causes B" if it is obtained at the expense of introducing arbitrary assumptions, which supposedly describe a situation that is absent in real economic environments? It seems that such a proof contributes nothing to identify what conditions should be found or created in *our world* to get B to guarantee F. We are in a situation that seems to be very close to that of the logical exercise outlined above.

In circumstances like these the technological moment comes to occupy the center stage in Alexandrova's account[8]. Starting from (2') that ensures that there are (unspecified) conditions out of which "F causes B", economists with practical orientation (and a host of other skillful people) can put their hands to work and try to find out concrete conditions C* (other than C), that can be implemented in our world and have the property that once imposed make F

---

[8] It is important to be clear about the particular type of laboratory experiment that concerns Alexandrova. Her analysis focused exclusively on the role of experimentation regarding *technological applications*, not for the purpose of testing models as this activity is usually understood. This is why she distinguishes between "test" and "testbed". The test of a model consists of creating or finding a situation in which model's assumptions are met, and see if their results are also obtained. In a testbed, on the other hand, it is known or supposed that the assumptions of the model are *not* satisfied. Its purpose is to obtain the same results obtained within the model from different or additional conditions than those referred in the model. Alexandrova's testbeds enhances the role of applied economics and the autonomy of the achieved results regarding the particular conditions described in models. Testbeds are better described as a practice performed by economic engineers than by experimental economists in the traditional sense of the term.

results in B. In fact, the main claim of Alexandrova is that this achievement has already been obtained with remarkable success in the case of the auction model. Maybe Alexandrova is right on this point, but since in her account models merely inspire (2') and do that in an extremely vague form, there remains the philosophical problem of assessing what exactly auction models' contribution is to the solution of the question raised by the hypothesis. Did the economic engineers referred to in Alexandrova's account need rigorous proof like the one provided by the model to find particular conditions C* under which doing F B is obtained in our world? Did they find, at least, a clue in the model to imagine the specific content of the set C*?

If "F causes B" is a desirable conclusion, there seem to be two different research programs concerning this result. One mathematical (logical): search for *any* conditions C under which the result could be deduced. Another, practical: find conditions C* feasibly implemented in our world such that the production of F leads to B. Unless a clear connection between both programs can be exhibited (something that Alexandrova's paper fails to show) to get busy in building models diverts resources from the technological approach of directly "building" in practice the desired result. This construction, it seems, does not need at all any of the solutions offered within the model.

Conclusions

The points of view examined in this paper agree in that actual conventional models that incorporate tractability assumptions provide *some* relevant information, but they must be supplemented with other types of knowledge, skills and practices if such information is to be successfully used in real world economies. Beyond this coincidence these views differ in the type of extra theoretical resources that are needed. For Cartwright models offer "lessons", which have to be extracted using pre-existing backward knowledge coming from outside the models themselves. Colander is more specific arguing that expert interpreters of current macroeconomic models are crucially needed, emphasizing the importance of having historical knowledge, methodological skills and experience in the analysis of particular situations. The economist-engineers portrayed in Alexandrova's account, on the other hand, are men of action. Thanks to the cooperation of other experts not necessarily economists, they can make – by trial and error-creative contributions to the design of institutions invested with economic relevance.

A major success of the perspective of Alexandrova, which in my opinion makes it superior to the rest of the views examined in this paper, is that she relates the epistemic relevance of economic models to their *practical applications*. From this point of view it is the social technology that the models help to generate which gives them credence as tools for achieving relevant knowledge. Indeed, if a discipline provides "resources" (models in this case) that contribute to successful technological devices (institutions, in this case), this is a clear indicator that this discipline brings out relevant and reliable knowledge (and some may feel entitled to apply to it the label of "science"). That is what has happened with physics, biology, and more recently with genetics. Also in the case of economics their practical applications are crucial, and so it is necessary to have a successful associated engineering. The problem is that, unlike what happens with the aforementioned disciplines that undoubtedly have contributed to an enormous amount of successful practical applications, the contribution of economic models to the generation of social technologies is still equivocal and needs to be properly examined. Indeed it is not clear whether there are or are not successful social

technologies. But even granting that they can be found in real economies, it remains unclear what exactly the contribution of models with arbitrary tractability assumptions has been in such cases. Philosophy of economics may be extremely helpful on this issue. One major contribution would be to shift philosophical attention from the intricate details of representations (models) to the conditions that have to be fulfilled for building a substantial core of successful applied economics.

Bibliography

Alexandrova, A. (2006), "Connecting Economic Models to the Real World: Game Theory and the FCC Spectrum Auctions", *Philosophy of the Social Sciences* 36: 173-192.

Alexandrova, A, (2008), *Making Models Count*, Philosophy of Science, 75, pp. 383–404.

Alexandrova, A, y Northcott, R (2008) "Progress in Economics: Lessons from the Spectrum Auctions" in Harold Kincaid and Don Ross (Eds.) *Handbook on the Philosophical Foundations of Economics as a Science* (Oxford: Oxford University Press, 2008).

Cartwright, N. (1999) "Vanity of Rigour in Economics" in *Discussion Paper Series*, Centre for the Philosophy of Natural and Social Science, LSE, 1-11.

Cartwright, N. (1995), "Ceteris Paribus Laws and Socio-Economic Machines", (Monist, July 95, Vol. 78, Issue 3)

Cartwright, N. (2007a), *Causal Powers: What Are They? Why Do We Need Them? What Can and Cannot Be Done with Them?*, CPNNS, LSE.

Cartwright, N. (2007b), *Hunting causes and using them*, Cambridge University Press

Cartwright, N. (2008) "Models vs Parables", Insights, Institute of Advanced Studies, Volume I, Number 11.

Colander, D., (2010), "The economics profession, the financial crisis, and method", Journal of Economic Methodology, Vol. 17, No. 4, December 2010, 419–427

Frigg, R. (2006) "Scientific Representation and the Semantic View of Theories", *Theoria*, 55: 49-65.

Giere, R., (1979), *Understanding Scientific Reasoning*, Holt, Reinhart and Winston, New York.

Giere, R., (2004), "How Models Are Used to Represent Reality", *Philosophy of Science*, 71, pp. 742 - 752.

Hausman, D.M., (1992), *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.

Knuuttila, T., (2004), "Models, Representation and Mediation", PSA.

Kuorikoski, J., Lehtinen, A. and Caterina Marchionni, C., 2010, "Economic Modelling as Robustness Analysis", Brit. J. Phil. Sci. 61 (2010), 541–567.

Kuorikoski, J., and Lehtinen, A., (2009), "Incredible Worlds, Credible Results", Erkenntnis, vol. 70, n° 1.

Lawson, T., (2003), *Reorienting Economics*, Routledge, London.

Lipsey, Richard & Steiner, Peter, (1981), *Economics*, Sixth Edition, Harper & Row, Publishers, NY.

Lucas, Robert E., Jr. 1981, *Studies in Business-Cycle Theory*, Cambridge, MA: The MIT Press.

Mäki, U. (1992) "On the Method of Idealization in Economics" *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26: 319-354.

Mäki, U., (ed.), 2002, *Fact and Fiction in Economics: Models, Realism, and Social Construction*. Cambridge: Cambridge University Press

Mäki, U., (2009), "Realistic realism about unrealistic models", en *The Oxford Handbook in the philosophy of economics*, ed. by Kincaid and Ross, Oxford University Press.

Morgan, M.S. and M. Morrison (1999), *Models as Mediators*, Cambridge: Cambridge, University Press.

Morrison, Margaret and Morgan, Mary, (1999), "Models as Mediating Instruments", in M. Morgan and M. Morrison (eds.), *Models as Mediators.* Cambridge: Cambridge University Press, 10-37.

Musgrave, Alan (1981): "'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted", *Kyklos,* vol. 34, no. 3, pp. 377-387.

Sugden, R. (2000) "Credible Worlds: the status of theoretical models in economics", *Journal of Economic Methodology* 7/1: 1-31.

Syll, L. P., (2010), "What is (wrong with) economic theory?, *real-world economics review,* issue no. 54

Weisberg, M. (2007), "Three Kinds of Idealization", *The Journal of Philosophy* 104, 12, 639-59.

Author contact: gustavoleomarques@hotmail.com

_____

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

A Copernican turn in Banking Union

Thomas Mayer[1]   [Goethe University and Deutsche Bank, Germany]

At every meeting of central bankers, policy-makers and economists, there seems to be agreement that creation of a "Banking Union" is essential for the survival of the euro. Yet, progress in building this union has been painfully slow. The Single Supervisory Mechanism (SSM) may not be ready before the middle of next year; the Single Resolution Mechanism (SRM) may end up as a complicated mixture between bail-in and bail-out instruments, funded at the national and euro-area levels; and common deposit insurance has been postponed into the indefinite future. What is making the establishment of Banking Union so difficult are the protracted fights over which government will be the payer of last resort when banks fail because of bad loans made in the past.

The recent agreement by the Eurogroup on the SRM highlights the dilemma: The European Stability Mechanism (ESM) can acquire stakes in banks in difficulties, but only up to a limit of €60 billion and with participation of the country in which the particular problem banks reside. Creditors other than holders of insured deposits are to be "bailed in" as a rule, but many exceptions to the rule are possible. Moreover, the European Commission's proposal for a bank restructuring and resolution regime is highly controversial and resisted by German officials and bankers who fear a transfer of national sovereignty to the Commission that is not backed by the EU Treaty.

If we continue along the present line, it does not seem likely that we shall ever reach full Banking Union. Therefore we need to learn from Copernicus who could not make sense of the movement of planets as long as he assumed that the sun moved around the earth. But once he assumed the opposite, everything fell into place for him. So, instead of trying to move from common bank supervision, over to bank resolution and then on to deposit insurance, let's go backwards and start with deposit insurance, move from there to resolution and end with supervision (see Table A1 in the Appendix).

Step 1. A 100% reserve requirement for safe deposits

We start by defining the risk-free asset for a euro-area resident with short-term and long-term financial liabilities (e.g. living expenses and nominal debt): This is the asset that can be converted into legal tender at face value at any time and under any circumstance. The concept of legal tender is very important in a fiat money system, in which money derives its value from government regulation or law, because it ensures that we can settle debt with almost worthless paper or electronic bits. In a fiat money system the only legal tender is by definition central bank money. Hence, an asset is risk-free if it can be converted into central bank money at any time. It is easy to see that only few assets would qualify as risk-free. Most importantly, the debt of governments that do not control the issuance of legal tender, as is the

---

case in EMU, or deposits of banks that are backed by credit to entities that also do not control the issuance of legal tender, are not risk-free. All these assets are risky because the debtor may not be able to convert them into legal tender at any time and under all circumstances.

Hence, in EMU, where governments have no access to the money printing press of the European Central Bank (ECB), the only risk-free asset is cash issued by the central bank and deposits that are fully backed by central bank reserves held at the central bank. No government sponsored insurance scheme can make deposits risk-free because this scheme cannot guarantee the availability of legal tender under any circumstances. It thus follows that we need to establish safe bank deposits as deposits that are fully backed by banks' holdings of central bank reserves. In other words, we can effectively insure deposits by introducing a 100% reserve requirement for this type of deposits. No industry or state deposit insurance scheme is required. A simple 100% reserve requirement is sufficient.[2]

But would a deposit insurance scheme based on a 100% reserve requirement be at all possible in our present system? The answer, of course, is yes: To back "insured" deposits created earlier by fractional reserve banking, banks could borrow central bank reserves in the necessary amount and keep them on deposit with the central bank. The cost of these safe deposits for the banks would be determined by the difference between the lending rate for central bank reserves and the deposit rate for central bank money. The cost for the bank customer would be determined by the net cost of central bank funds for the banks and the banks' operating costs for the insured deposits. The benefit for the customer would be to have a safe asset other than only central bank notes, and the ability to use this asset to make non-cash payments.

A quantitative limit for safe deposits would not be necessary as the central bank could adjust the supply of reserves to the demand for safe deposits. But the central bank could influence the demand for safe deposits by changing the variable costs, which are given by the difference between the cost of central bank reserves and the rate that the central bank pays on deposits. This could be used for stabilisation policy: By influencing the demand for safe deposits relative to other deposits, the ECB would also influence credit extension by the banks.

Assume that customers switch from investor deposits to safe deposits. If the ECB kept the supply of central bank reserves constant, banks would need to reduce credit to free funds for deposit with the ECB as cover for the additional safe deposits. Credit to the non-bank sector would go down, and the credit multiplier, defined as credit relative to central bank money, would fall. Alternatively, if the ECB wanted to accommodate the switch and keep credit to the non-bank sector constant, they could increase the supply by central bank reserves to meet the additional demand. Still the credit multiplier would decline, albeit by less than before, because the central bank money stock would increase. Finally, if the ECB wanted the credit multiplier to remain constant, they could raise the alternative costs of holding safe deposits by lowering the deposit rate. The reduction of the deposit rate needed to achieve the target level

---

[2] The idea of 100% reserve coverage of deposits is of course not new. As Huerta de Soto has pointed out, it dates back to the school of Salamanca in the 16th century, was taken up in the UK Bank Charter Act of 1844 and is advocated today by followers of the Austrian School of Economics and others (see Jesus Huerta de Soto, "Money, Bank Credit, and Economic Cycles", Ludwig von Mises Institute, Auburn AL, 2012).

of safe deposits could be determined in a reverse refinancing operation, where banks submit bids for the deposit rate they are willing to accept (or pay when the deposit rate is negative).[3]

Step 2. A hierarchy of loss-absorbing bank liabilities

Once we have established reserve-backed deposits as safe assets, all other bank liabilities would of course be risky. We can now define a hierarchy of loss absorption in a bank resolution regime. The first loss would be borne by the equity tranche on the liability side of banks' balance sheets. After having set aside assets pledged to cover secured debt, the second and third losses would be borne by junior and senior unsecured bank debt. The fourth and last loss would accrue to deposits uncovered by central bank reserves. When all bank liabilities except deposits fully covered by central bank reserves contribute to cover losses on bank assets, taxpayer-funded bank bailouts would become significantly less likely (and will eventually become unnecessary). As long as banks engage in maturity transformation, systemic liquidity crises remain possible and a lender-of-last resort necessary. But the risk of liquidity support turning into support for insolvent banks would be much diminished when there is a clear roadmap for bank resolution. Moreover, the risk of a liquidity crisis could be reduced if the scope for maturity transformation would be limited in the regulatory framework. Finally, when the public fully understands the risk associated with an exposure to banks beyond the reserve-backed safe deposit, it would be up to banks to reassure bank equity investors and creditors that their assets are being managed in a way that makes illiquidity and losses become unlikely.

Step 3. Divest banks from governments by revised regulations for government debt

To be able to fund their assets at reasonable costs, banks would need to have a comfortable equity cushion and a well-diversified and reasonably liquid portfolio of assets. Most importantly, they would have to reduce their exposure to government debt to a level consistent with this debt being subject to default risk. Hence, in the new regulatory regime, government debt would have to be backed by equity and other loss-absorbing bank liabilities, and it would have to be subject to limits for single credit exposure. To allow banks' divestment from government debt, the European Central Bank could buy in a one-off operation the government bonds that banks have pledged to the central bank as collateral for obtaining central bank credit, and place them in a special account that will be wound down over time.[4] As a result of this operation, risky claims of the banks on governments would be replaced by risk-free claims of the banks on the ECB or, in other words, by central bank reserves. The ECB would of course want to reduce its exposure to government debt over time.

Since it is very doubtful that all highly indebted euro-area countries could repay their debt, governments and the ECB could agree that all income from seigniorage would be used to pay down the government debt held by the ECB in the special account. In practice this would

---

[3] Banks in Germany and certain other euro area countries today already hold large amounts of central bank reserves. However, these reserve holdings are motivated by the banks' reluctance to lend to other banks in other euro area countries and are not earmarked to back deposits.

[4] Since government debt presently does not need to be backed by bank equity, is not subject to single credit exposure limits, and is liberally accepted by the ECB as collateral for loans of central bank money to banks, banks have in effect become intermediaries for ECB credit to governments (though this has been camouflaged by keeping government credit on banks' balance sheets). To end this practice, the ECB will have to properly account for its true exposure to government debt by assuming the credit to government presently parked on banks' balance sheets on its own balance sheet.

mean that the ECB instead of governments would redeem maturing (or repurchase outstanding) bonds and debit governments' seigniorage account with the costs of the transaction. Since the present discounted value of seigniorage can be very large, reaching several trillion euros in the case of the euro area, depending on interest rates on central bank credit and the growth rate of non-interest-bearing central bank money, it seems likely that this would be sufficient to eventually retire the government debt acquired by the ECB from the banks. The arrangement outlined here has some resemblance to the debt redemption fund proposed by the German Council of Economic Experts. However, an important difference is that in the arrangement proposed above, the ECB would withhold revenue to pay down the debt and would not have to rely on governments to allocate revenue for this purpose.

Part of the reserves obtained by selling government bond holdings to the ECB could be used initially by the banks to back safe deposits. The rest could be released by the ECB into the banking system and the economy at large by setting a rate for central bank deposits below the risk-adjusted bank lending rates. With their debt now subject to default risk, highly indebted governments might encounter difficulties accessing the market at reasonable costs to roll over expiring debt. But market access could be improved if the ECB agreed to assume the status of a junior creditor for the government bonds they have acquired from banks in case of a debt restructuring. Like the orderly pay down of the debt, the costs for such a restructuring could be covered by future seigniorage income. This would represent a partial mutualisation of public debt, but because of its limited character it would probably be acceptable for countries with stronger balance sheets.

Balance sheet effects of safe deposits

Table A2 in the Appendix shows the structure of banks' balance sheets after the introduction of safe deposits. Abstracting from assets earmarked for covered bonds, banks would have central bank reserves and credit on the asset side of the balance sheet, as before. However, central bank reserves would be tied to cover safe deposits on the liability side of the balance sheet. All other liabilities would participate in loss absorption in a clearly defined hierarchy, with equity providing the first layer and investor deposits (not covered to 100% by central bank reserves) the last.

Given our definition of a safe deposit, it corresponds to what are at present called "sight deposits". In April 2013, sight deposits in the euro area amounted to €4.4 trillion, representing about 38% of total deposits or 44% of GDP. Since customers would probably not choose to have all sight deposits in the form of safe deposits, this would represent an upper boundary to the level of safe deposits. In April 2013, banks held €556 billion as reserves with the Eurosystem (€273 billion of which counted as minimum reserves). Hence, the introduction of safe deposits would substantially increase reserve holdings and the Eurosystem's balance sheet (presently only €2.5 trillion). But this would only change the mix between inside and outside money and not affect the overall size of the balance sheet of the monetary and financial system.

Table A3 in the Appendix shows the structure of the balance sheet of the ECB. As can be seen from this table together with Table A2, safe deposits, like bank notes in circulation, represent a direct liability of the ECB to the non-banking sector. Against this stands the 'good will' on the asset side of the central bank's balance sheet, which reflects the trust invested by the public in money as a means of exchange and store of value. At first glance, the backing of

money by "good will" in the central bank's balance sheet may look unsound. Proponents of 100% reserve backing of deposits have therefore suggested that the central bank acquire government debt to back safe deposits and issue money against new government bonds when it wants to increase the central bank money stock. But this only camouflages the lack of a material cover of money in a fiat money system and creates circularity in the accounts: The claim of the central bank on the government is neutralized by the government's eventual authority over the central bank. Because of this the government may be tempted to use direct central bank purchases of government debt as an excuse to fund its expenses through the money printing press. The fact is that the only cover of money in a fiat money system is people's trust in money, and this is most honestly accounted for by 'good will' in the central bank's balance sheet.

As explained above, the central bank can influence the mix between safe deposits and investor deposits by determining the alternative costs of safe deposits. Since investor deposits fund bank credit, this allows the central bank to influence credit extension by the banking system. Banks can of course still engage in maturity transformation by funding longer-term credit with rolling short-term investor deposits. But holders of investor deposits would be exposed not only to credit but also to liquidity risks associated with maturity transformation. Since they would demand a risk premium as compensation, there would be an economic limit to maturity transformation.

In a growing economy, the central bank may not only want to influence the mix between safe and investor deposits but also the size of the balance sheet of the banking sector. It can do so by writing up 'good will' in its balance sheet and crediting safe deposits with this amount (i.e., paying safe deposit holders something like a dividend). A write-up of 'good will' could be triggered by an increase in the demand for money as a result of an increase in potential GDP. In this case, the price level would fall if no new money was issued. As long as price rigidities exist, this may not seem desirable. Thus, new central bank money would come into existence in a neutral way and would not benefit any sector in particular (as would be the case if the central bank would create new money by buying newly issued government bonds, as suggested by some). Economists of the Austrian school have pointed out that the creation of money via bank credit or government spending benefits those close to the process of money creation and puts at a disadvantage those far away from it. The latter will not obtain new money but may suffer from price increases triggered by the money injection.

A more level playing field

The proposed structure for Banking Union would of course change the way in which banks operate and governments fund themselves. Banks would no longer extend credit and create book money at will. Rather, they would assume the dual role of 1) safe keeper of the risk-free assets, i.e. central bank money, for depositor-savers, and 2) intermediary of funds between investor-savers and entrepreneurs. There would be no need to limit "deposit insurance" to a certain amount, e.g., the EUR100,000 now in force in the euro area, and deny larger depositors, e.g. companies, access to a safe store of value. As safe keepers of the risk free asset and facilitators of non-cash payments banks would of course be entitled to a fee for the services they render, which would become a permanent source of revenue for them. By the same token, banks would receive fees for acting as agents in capital markets and generate income as risk takers when using investor-deposits as a source for credit.

It is possible that bank lending rates would increase in the new regime, but if they do, then only because savers realise that in a fractional reserve banking system bank deposits carry credit risk, unless they are fully backed by banks' holdings of central bank reserves. In fact, the widespread misconception that bank deposits in our present system of fractional reserve banking are completely safe and can be converted into central bank money at any time and in all circumstances represents a subsidy to bank lending rates (and bank profits) from tax payers, who in times of crises are called upon to stabilize banks.

Governments could no longer rely on banks to fund their debt and would have to obtain funding from the capital markets. Borrowing costs could also increase for them as they would no longer be regarded as offering risk-free assets and could no longer benefit from preferential treatment on banks' balance sheets in the form of zero-risk weighting for the calculation of regulatory capital requirements and exemption from single-credit exposure limits. Again, such an increase in borrowing costs would represent the end of a subsidy to government borrowing as a result of special regulatory treatment.

Conclusion

To sum up, Banking Union could be built in three steps. In the first step, "deposit insurance" could be introduced in the euro area by requiring banks to fully back safe deposits with central bank reserves. This would be the only safe asset in EMU, where, as already noted, governments have no command over the money printing press of the central bank. All other bank liabilities would participate in covering losses on the asset side of banks' balance sheets in a hierarchical order established by the common bank resolution regime in the second step. To help banks divest from government bonds, the ECB could buy these bonds from them, replacing risky claims of banks on governments by risk-free claims of banks on the ECB in the third step. Governments and the ECB could agree to use future seigniorage income to pay down the government debt held by the ECB.

Appendix

Table A1. The Copernican turn for Banking Union

Present approach	Proposed approach
Step 1 Establish SSM on the basis of the regulatory framework mapped out in CRDIV.*	**Step 1** Establish deposit insurance by requiring safe deposits to be backed 100% by banks' holding of reserves with the central bank.
Step 2 Establish SRM backed by a government-funded restructuring and resolution fund.	**Step 2** Establish SRM with hierarchical loss absorption of all bank liabilities except safe deposits. Resolution fund would operate only in the transition to new regime, and then would no longer be required.
Step 3 Keep deposit insurance under national authority.	**Step 3** Establish SSM on the basis of CRDIV, modified to introduce positive risk weights and single credit exposure limits for government debt. The ECB would help divest banks from government bonds and redeem the ECB's acquired government bond portfolio by withholding seigniorage income over time.

* CRDIV is the legislative package of proposals adopted by the European Commission on 20 July 2011 to replace the current Capital Requirements Directives and strengthen the regulation of the banking sector.

Table A2. The structure of bank balance sheets in the new regime

Assets	Liabilities
Central bank reserves	Safe deposits
Ring-fenced assets	Covered bonds
Other assets	Investor-deposits* Senior debt* Junior debt* Equity*

* Participating in losses in ascending order.

Table A3. The structure of the central bank's balance sheet in the new regime

Assets	Liabilities
Good will	Deposits of commercial bank reserves to cover safe deposits
Other assets	Other liabilities, reserves, and capital

Author contact: tom.mayer@db.co

_____

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

A monetary and fiscal framework for macroeconomic stability in the European Monetary Union

Thomas Oechsle[1] [KPMG, Germany]

Abstract
A proposal for a monetary and fiscal framework for the general case, i.e economies outside the eurozone, is set out. After examining the distinctive features of the eurozone, the paper proposes a new monetary and fiscal framework for the European Monetary Union. The monetary policy regime proposed for the eurozone is the same as in the general case. However, it is argued that, with respect to fiscal policy, a deviation from the first-best solution (to be implemented in the general case) would be beneficial in the case of the European Monetary Union.

Keywords: Monetary Policy, Fiscal Policy, Macroeconomic Stabilization, Optimum Currency Area Theory

1 Introduction

In the case of the European Monetary Union (EMU), the conditions for an optimal currency area are far more starkly violated than, say, in the USA. Mundell (1961) pointed to factor mobility as the single most important determinant of the workability of a single currency area. But labour mobility between member states is heavily constrained due mainly to language barriers.

The basic requirement for the workability of a currency union with heavily constrained labour mobility between its member states (such as the EMU) is that the price levels in all member countries rise by the same proportion over time.

Substantial differences in the price levels of the regions of a currency area are, of course, always undesirable. The question is only whether it would be beneficial to accept other distortions in order to avoid the possibility of the emergence of substantial differences in regional price levels. The answer to this question depends, of course, on how costly the elimination of large regional price levels would be, should they arise.

The ongoing eurozone crisis shows that the economic and social costs associated with realigning the regional price levels in the EMU are tremendous. There is no good reason to assume that the costs associated with the removal of significant regional differences in competitiveness would be significantly lower in the foreseeable future.

This paper argues that it would therefore be desirable to accept the (limited) distortions/costs associated with deviating from the first-best solution for a macroeconomic framework in order to avoid the potentially huge costs associated with the possible emergence of substantial differences in regional price levels within the eurozone.

---

[1] Associate at KPMG Life Member of Clare Hall, University of Cambridge

Section 2 sets out a proposal for a monetary and fiscal framework for the general case (i.e economies outside the eurozone). The monetary framework proposed is described in subsection 2.1, while the optimal fiscal framework is outlined in subsection 2.2.

2   The general case

2.1   A proposal for a monetary framework: "Krugmanite" NGDPRT

2.1.1   What is "Krugmanite" NGDPRT?
With "Krugmanite" Nominal GDP Rate Targeting (NGDPRT) I mean a monetary policy regime under which the central bank has the mandate to target some growth rate of NGDP, say, 5%. The central bank attempts to steer the actual growth rate of NGDP towards the target through the use of interest rate changes and other monetary tools. If these tools are not enough to meet the target rate of NGDP growth, the central bank commits to higher future NGDP growth in order to achieve the NGDP growth target today.[2]

Assume, for example, that the target growth rate of NGDP is 5%. Then the economy is hit by a strong aggregate demand shock. Trying to offset the shock, the central bank cuts interest rates. However, even after cutting interest rates all the way to zero, market expectations indicate that the NGDP growth target will be missed.[3] The central bank is therefore obliged to commit to higher future NGDP growth in order to meet the target today. It may, for example, commit to target 6% rather than 5% NGDP growth next year.

If such a commitment is not enough to meet the target today, the central bank commits to even higher future NGDP growth, say, 7% next year or 6% for the next two years.

This procedure may be repeated several times until market expectations indicate that the central bank will meet its target in this period.

2.1.2   Why "Krugmanite" NGDPRT?
I will answer this question in two parts:
>   1. Why NGDP Rate Targeting rather than inflation targeting?
>   2. Why "Krugmanite" NGDP Rate Targeting rather than "normal" NGDP Rate Targeting or NGDP Level Targeting?

One advantage of NGDP Rate Targeting over inflation targeting is that NGDPRT does not exacerbate the effect on output of negative supply shocks. Another advantage put forward by proponents of NGDPRT is that, following a positive productivity shock, NGDPRT would be less conducive to the creation of asset price bubbles than inflation targeting.

However, by far the most important advantage of NGDPRT over inflation targeting pertains to aggregate demand shocks.

---

[2] Krugman (1998) was the first to explicitly point out that, even if (due to the zero lower bound) conventional monetary policy has lost traction, a commitment to a more expansionary monetary policy in the future can increase aggregate demand today. Hence, I use the term "Krugmanite" NGDPRT to describe the monetary framework proposed. Other important contributions to monetary policy at the zero lower bound include Eggertsson and Woodford (2003) and Jung, Teranishi and Watanabe (2005).
[3] Obviously, at the moment, it is not possible to know exactly what the markets' expectations of NGDP growth are. A NGDP futures market, as proposed by Scott Sumner (see, for example, Sumner, 2011b) would provide the central bank with the relevant information.

Under the assumption of reasonably small lags between changes in production on the one hand and changes in nominal wages (and prices) on the other, inflation targeting would do a reasonably good job offsetting aggregate demand shocks: if output fell below potential, inflation would fall below its target. By loosening monetary policy, the central bank could bring inflation back to the target and output back to its potential level.

But if there is substantial nominal wage stickiness, inflation targeting is associated with severe problems. Consider an economy with a natural growth rate of 3% and an inflation target of 2%. Assume that this economy is hit by an aggregate demand shock so strong that, ceteris paribus (i.e. under the assumption that there is no responsiveness of nominal wages/prices or monetary policy to the shock), the real growth rate of the economy would fall to 0% (i.e. output would be 3% below potential).

In order to understand the failure of inflation targeting under substantial downward rigidity of nominal wages and prices, it is best to use the assumption of zero sensitivity of the rate of inflation to downward changes in the real growth rate as a reference point.

In this case the aggregate demand shock mentioned above would (in the absence of action by the central bank) reduce the real growth rate of the economy to 0% but inflation would not deviate from its target level of 2%. If the central bank could easily distinguish between offsettable (aggregate demand) shocks and non-offsettable (supply) shocks, everything would still be fine. The central bank would recognize the shock as an aggregate demand shock and offset it by loosening monetary policy. Output would not fall below its potential.

But now assume that the central bank cannot distinguish between aggregate demand shocks and supply shocks. In this case, there would be no compelling reason for it to loosen monetary policy. The real growth rate of the economy may have dropped to 0% because of an aggregate demand shock or as a result of a supply shock. In the latter case more expansionary monetary policy would merely lead to overshooting inflation. By not doing anything the central bank can at least ensure that it will meet the inflation target, i.e. its mandate.

In the real world recognizing aggregate demand shocks as aggregate demand shocks (and supply shocks as supply shocks) seems to be quite difficult for central banks (as well as other policymakers). Throughout the last years macroeconomists across the Western world have been occupied with arguing about the gap between the natural and the actual level of output - whether there is such a gap and what its size might possibly be.

And there has been a wide range of answers to these questions. Some central bankers and economists estimated the gap between the natural and the actual level of output to be hugely negative, while others argued that there was no deviation of output from its natural level at all. Some even suggested that output was slightly above its natural level for certain countries (such as the UK where inflation has been running above target for several years). Correspondingly policy recommendations have ranged from urgent calls for aggressive monetary stimulus to demands for tighter money.

In summary, inflation targeting would do a good job offsetting aggregate demand shocks if there were no wage and price rigidities or if the central bank had complete information. In

reality, there is substantial downward wage and price rigidity and the central bank does not have complete information.

Under NGDPRT, the central bank would automatically offset aggregate demand shocks – without having to have detailed information about the output gap (or the ability to distinguish between aggregate demand and supply shocks) and independent of the degree of wage/price rigidity.

Of course, conventional monetary policy may not be enough to fully offset an aggregate demand shock. In order to still meet the target growth rate of NGDP, there are two options available: fiscal policy or committing to a more expansionary monetary policy in the future.

Given the constraint on conventional monetary policy through the zero lower bound of nominal interest rates, macroeconomic stabilization could rest on a two-tier approach such as this: as long as the economy is not at the zero lower bound on nominal interest rates, the central bank uses conventional monetary policy to meet its mandate. If the economy is up against the zero lower bound, the central bank is empowered to determine the budget balance (the difference between government spending and tax revenues) so as to meet its mandate.

However, such a two-tier approach to macroeconomic stabilization would be suboptimal. Ideally, macroeconomic stabilization policy does not only produce the level of GDP (employment) that would be achieved under flexible prices. It also produces the allocation of resources achieved under flexible prices. Using fiscal policy as a macroeconomic stabilization tool can yield the level of GDP corresponding to the flexible price equilibrium but, obviously, it is not able to achieve the same composition of it.

Using fiscal policy as a counter-cyclical stabilization tool involves a trade-off between the goal of macroeconomic stabilization and the purpose of microeconomic efficiency. This kind of trade-off does not exist in the case of monetary policy, which therefore dominates fiscal policy as a macroeconomic stabilization tool.

This dominance also holds when conventional monetary policy is constrained by the zero lower bound and the central bank has to commit to future monetary policy actions in order to increase aggregate demand today (see, for example, Mankiw and Weinzierl, 2011).

Given that unconventional monetary policy in the form of a commitment to higher inflation (NGDP growth) in the future should produce less distortion than fiscal stimulus, the monetary framework should ideally enable the central bank to commit to a more expansionary future monetary policy in case the economy is up against the zero lower bound.

One way of doing this would be NGDP Level Targeting (NGDPLT).[4] If, under an NGDPLT regime, the target path of NGDP is undershot, this automatically leads to anticipation of a (temporarily) more expansionary monetary policy in the future. Expectations of monetary expansion in the future will reduce the real interest rate and therefore increase aggregate demand today, which will limit the degree of the undershooting.

---

[4] For a detailed description of the concept of NGDP Level Targeting see, for example, Sumner (2011a, 2011b).

The problem with NGDPLT is that, for expectations of higher NGDP growth in the future to materialize, the growth rate of NGDP today has to fall below its (implicit) target level. That is, NGDPLT implies a commitment to more expansionary monetary policy in the future only if the (implicit) target growth rate of NGDP (say 5%) is undershot today.

Hence, if the economy is hit by an aggregate demand shock that (due to the zero lower bound of nominal interest rates) cannot be fully offset by conventional monetary policy, NGDPLT can limit the degree to which the growth rate of NGDP undershoots the target. But monetary policy will not be able to prevent the growth rate of NGDP from undershooting the target rate. Obviously, it is not ideal for a monetary policy regime to imply a commitment to more expansionary monetary policy in the future merely as a reaction to NGDP undershooting its target growth rate. Rather, it would be desirable that such a commitment take place in order to avoid the growth rate of NGDP to be undershot in the first place.

And this is where the advantage of "Krugmanite" NGDPRT over NGDPLT lies. "Krugmanite" NGDPRT means that the central bank commits to a more expansionary monetary policy in the future, as soon as market expectations indicate that, despite zero interest rates, in the absence of further action the NGDP growth target will be undershot. That is, in contrast to NGDPLT, "Krugmanite" NGDPRT fully offsets *every* aggregate demand shock - independent of its size.[5]

The key to understanding why the central bank (provided it were endowed with the appropriate mandate) would be able to offset every aggregate demand shock is the simple fact that it can always print more money.

Of course, just increasing the money supply is not enough. In order for aggregate demand to increase, markets have to believe the increase in the money supply will be permanent. A permanent increase in the money supply will eventually lead to a proportional rise in the price level. At the zero lower bound, higher expected prices reduce the real interest rate. Hence, aggregate demand increases today.

Since the central bank can always print more money, there is no limit to the extent to which it can increase the future price level. Provided that markets expect the increase in the money supply to be permanent, the central bank can reduce the real interest rate up to the point where any aggregate demand shock is fully offset.

Committing to higher NGDP growth in the future is a means to convey to markets the information that the increase in the money supply will indeed be permanent.

2.2 The optimal fiscal framework

With respect to the expenditure side of fiscal policy optimality implies that the marginal benefit of government consumption be equal to the marginal benefit of private spending. That is, government consumption should remain a constant fraction of GDP over the business cycle. With respect to taxation optimality implies that tax rates are to remain constant over the

---

[5] Under NGDPLT, an aggregate demand shock which cannot be fully offset by conventional monetary policy causes the growth rate of NGDP to undershoot the target today and then overshoot the target in the future. Under "Krugmanite" NGDPRT, an aggregate demand shock which cannot be fully offset by conventional monetary policy does not cause the growth rate of NGDP to undershoot the target today; it only leads to above target growth of NGDP in the future.

business cycle in order to minimize the excess burden of taxes (see, for example, Barro, 1979).

Due to the counter-cyclical nature of transfer payments (automatic stabilizers) optimal fiscal policy would thus correspond to a counter-cyclical pattern in the budget balance (see, for example, Alesina and Tabellini, 2005).

However, one cannot plausibly explain the persistence of large deficits and the corresponding explosion of public debt observed in many countries as the result of optimal fiscal policy. At the root of this deficit bias are political economy distortions.

Political economy models of fiscal policy can be grouped into differential categories depending on the source of interest heterogeneity causing fiscal deficits (see Eslava, 2010):

> 1. heterogeneity of interests between politicians and voters
> 2. heterogeneity of interests between politicians
> 3. heterogeneity of interests between social groups or districts/regions.

In the first of these categories opportunistic policymakers use deficits to enhance their chances of getting re-elected. The second category of models is based on the hypothesis that politically motivated deficits are caused by partisan policymakers, who use fiscal deficits strategically to influence the policy of successors belonging to the opposing party. The third category explains the deficit bias as a result of distributional conflicts between social groups, districts or regions.

However, the deficit bias of the political sphere can be eliminated by a fiscal rule requiring the structural budget balance to be zero at all time, whereby estimates of the structural budget balance should be conducted by an independent institution. And this is exactly the fiscal framework for the general case proposed in this paper.

3   A framework for macroeconomic stability in the EMU

Section 2 made a proposal for a monetary and fiscal framework for the general case, that is, for currency areas that are either optimal or not suboptimal enough to justify a deviation from the first-best solution for the monetary and fiscal framework.

The theory of the second best states that, if one optimality condition cannot be satisfied, the next-best solution involves changing other conditions away from optimality as well (see Lipsey and Lancaster, 1956).

In practice, policymakers usually lack the information to implement the second best policies. According to the theory of the third best (see Ng, 1983, chapter 9), policymakers should therefore refrain from trying to reduce the cost associated with one kind of distortion by artificially introducing other distortions into the economy.

For this reason, I do not suggest that, say, for the USA, modifying the monetary and fiscal framework described in section 2 would be preferable to implementing the first-best solution - even though I agree with Kouparitsas (2001) that the US is not an optimum currency area.

However, in the case of the European Monetary Union (EMU), the conditions for an optimal currency area are far more starkly violated than in the USA. Mundell (1961) pointed to factor mobility as the single most important determinant of the workability of a single currency area. While capital is certainly sufficiently mobile in the EMU, this is not at all the case for labour. Labour mobility between member states is heavily constrained due mainly to language barriers.[6]

The economic and social costs resulting from asymmetric shocks in combination with immobility of labour have turned out to be extremely large. Given that there is no reason to believe that, in the future, the EMU will not be subject to (large) asymmetric shocks anymore and given that the immobility of labour between the member states is unlikely to decrease significantly in the foreseeable future, a deviation from the first-best solution for the monetary and fiscal framework is, in my view, justified in the case of the eurozone countries.[7]

The introduction of the Euro reduced real interest rates in the peripheral countries of the eurozone, which led to a significant increase in aggregate demand. The development in Germany was the exact opposite: investment was sluggish for years and unemployment was high. As a result, the peripheral countries were consistently overshooting the common inflation target. Germany, on the other hand, was consistently undershooting the inflation target.

The peripheral countries with their relatively fast rising price levels became increasingly uncompetitive vis-à-vis Germany with its relatively slowly rising price level. When private capital flows from Germany to the peripheral countries came to a sudden end, the peripheral economies were left with prices substantially out of line with those in Germany.

The cost associated with the internal devaluation the peripheral countries are now going through would be greatly mitigated if there were high labour mobility between the member countries of the eurozone. But due mainly to language barriers, workers from, say, Spain cannot easily emigrate and take up jobs in, say, Germany.

The basic requirement for the workability of a currency union with heavily constrained labour mobility between its member states (such as the EMU) is therefore that the price levels in all member countries rise by the same proportion over time.[8] Put differently: while a member state of the EMU may no longer have its own monetary policy, it still has an "inflation target" to meet year by year.

But how can each member state of a single currency area ensure to meet the common "inflation target" without having its own monetary policy? The answer is: through using fiscal policy.

---

[6] Another important determinant of the workability of a single currency area is the degree of fiscal integration (see Kenen, 1969). Since the EMU is not a federal state but an association of independent nation states, fiscal integration within the EMU is virtually non-existent.

[7] In the long run, labour may become much more mobile within the EMU, if the English language in Europe evolves to being used for business and communication in way it is already used in, say, India today.

[8] Of course, avoiding substantial differences in the price levels of the regions of a currency area would also be desirable if labour were highly mobile. But if labour is as immobile as in the EMU, avoiding substantial differences is absolutely crucial: without the mitigating effect of high labour mobility, the costs associated with realigning the price levels of different regions are extremely high.

Given the fact that member countries of the EMU are subject to asymmetric shocks and given the fact that they have transferred monetary policy to a common central bank, fiscal policy becomes crucial for the purpose of stabilizing the national economies and ensuring that price levels in all member countries rise by the same proportion over time.

As described above, the introduction of the Euro led to lower real interest rates and an economic boom in the periphery, while the development in Germany was the exact opposite. Here is what should have happened: fiscal policy in peripheral countries should have been very contractionary throughout the boom years in order to meet the common inflation target. For the same reason (namely to achieve the common inflation target) fiscal policy in Germany should have been relatively expansionary. This would have stimulated demand, output and therefore wages.

Here is what actually happened: in the peripheral countries fiscal policy was far too expansive given the economic environment. As long as the deficit limit of 3% – erected by the "Stability and Growth Pact" (SGP) – was not exceeded, governments in those countries did not have sufficient incentives to tighten fiscal policy in order to meet the common inflation target. As a result, those countries were consistently overshooting the common inflation target.

Germany, on the other hand, was consistently undershooting the inflation target: instead of being expansive, fiscal policy in Germany, in its attempt to meet the requirements of the SGP, was contractionary, thereby hampering demand (and therefore output and nominal wage increases) even more.

That is, rather than offsetting the asymmetric shocks hitting the eurozone, fiscal policy in the member countries of the EMU exacerbated them by becoming pro-cyclical.

Neutral fiscal policy[9] (i.e. the first-best solution for a fiscal framework, see subsection 2.2) would have been an improvement on the pro-cyclical fiscal policy observed in the eurozone but it would most probably not have prevented the emergence of substantial differences between regional price levels.

Substantial differences in the price levels of the regions of a currency area are, of course, always undesirable. The question is only whether it would be beneficial to accept other distortions (i.e. costs) in order to avoid the possibility of the emergence of substantial differences in regional price levels. The answer to this question depends, of course, on how costly the elimination of large regional price levels would be, should they arise.

The ongoing eurozone crisis shows that the economic and social costs associated with realigning the regional price levels in the EMU are tremendous. There is no good reason to assume that the costs associated with the removal of significant regional differences in competitiveness would be significantly lower in the foreseeable future.

In my view, it would therefore be desirable to accept the (limited) distortions/costs associated with deviating from the first-best solution for a fiscal framework in order to avoid the potentially huge costs associated with the possible emergence of substantial differences in regional price levels within the eurozone.

---

[9] Note that "neutral" fiscal policy would – due to the counter-cyclical nature of transfer payments (automatic stabilizers) – correspond to a counter-cyclical pattern in the budget balance (see subsection 2.2).

Each member country of the EMU should be subject to a fiscal rule requiring the budget balance to be set in such a way as to ensure that, over time, the national price level increases by the same proportion as the price level in the eurozone as a whole. That is, on the national level there should be a price level targeting mandate for fiscal policy. Such a fiscal framework would have prevented the inflation/competitiveness divergence between Germany and the peripheral countries of the EMU.[10]

The ECB's monetary policy regime (whether it be inflation targeting, price level targeting, NGDPRT, NGDPLT, "Krugmanite" NGDPRT or yet some other policy regime) determines how the "common" (eurozone-wide) price level evolves over time.

A price level targeting mandate for fiscal policy implies a commitment of member states to correct past deviations from the eurozone-wide change in prices and would thus ensure that national price levels rise roughly synchronously over time. If, in a given year, inflation in the eurozone as a whole were, say 2%, but inflation in, say, Spain were 3%, then Spain would have to correct for this deviation from the eurozone-wide rate of inflation by targeting a rate of inflation below the eurozone-wide average in the next year.

In any given year, the expected rate of inflation for the eurozone as a whole is the reference point that national fiscal authorities use in order to determine which rate of inflation to aim for in that year. For a member state whose price level has risen synchronously with the eurozone-wide price level over time, the inflation target for that year will, of course, simply equal the expected eurozone-wide rate of inflation.

Price level targeting (through the use of fiscal policy) on the national level does not imply constraints on the choice of the monetary policy regime on the European level. If the ECB followed an NGDP (Rate or Level) Targeting regime, the expected rate of inflation for the eurozone as a whole (i.e. the reference point for national fiscal authorities) would simply equal the difference between the NGDP growth target and the expected real growth rate.

That is, a deviation from the first-best solution for the fiscal framework on the national level does not necessitate a deviation from the first-best solution for the monetary framework on the European level. Hence, "Krugmanite" NGDPRT could and, in my view, should be the monetary policy regime of the ECB.

Some might argue that, due to political economy distortions, fiscal policy could not be relied upon to fulfil its price level targeting mandate.

Discretionary fiscal policy has undoubtedly been subject to a deficit bias across countries. However, under the fiscal framework proposed in this section, fiscal policy would not be discretionary but subject to a clear mandate. And as long as the sanctions on a government not meeting its mandate are sufficiently costly, the framework should provide the member countries of the eurozone with the right incentives to meet their price level targeting mandates.

---

[10] If, at some point in the future, labour is much more mobile within the EMU (for example because everybody is fluent in English), there is, of course, no justification any more for deviating from the first-best solution for a fiscal framework. Fiscal policy should then be required to be neutral.

Alternatively, the determination of the budget balance (i.e. the difference between public spending and tax revenues) could be delegated to an independent institution: independent institutions such as the central bank are not subject to the political economy distortions associated with the democratic sphere.[11]

In each member country of the eurozone a *National Fiscal Policy Committee* could be given the mandate to set the budget balance in such a way as to ensure that, over time, the national price level increases by the same proportion as the price level in the eurozone as a whole.[12] Such *National Fiscal Policy Committees,* independent of politics and given a clear mandate, would provide the right incentives for fiscal policymakers.

Of course, delegating the determination of the budget balance target to an independent institution would be associated with one potential problem. The committee sets the budget balance target having in mind a certain aspired effect on aggregate demand (and hence the rate of inflation).

But different expenditures have different multipliers and the same is true for different taxes. Different compositions of tax and expenditure changes consistent with the committee's budget balance target may have significantly different effects on aggregate demand (Calmfors, 2003, p. 336). Therefore, the actual effect of the budget balance target on aggregate demand may be quite different from the one the committee had in mind when setting the target.

Fortunately, there is a simple solution to this problem: the committee should have the right to adjust the budget balance target at any time. If it transpires that the actual effect of the fiscal change differs from the intended effect on aggregate demand, then the committee can adjust the budget balance target accordingly.

One might think that delegating the determination of the budget balance to an independent institution would not be democratically legitimate. But this is not the case.

Generally, delegation of decision-making powers to an independent institution is innately democratically legitimate, if any disagreement regarding the question of how to best achieve a certain objective (which has to be either a common objective shared by virtually everybody or one that has been determined democratically) is not caused by different value judgements but only by different factual judgements (see, for example, Calmfors, 2003, pp. 334 - 335).

The determination of the budget balance does not involve value judgements because it does not involve non-negligible redistributive effects. Delegating the determination of the budget balance to an independent institution, while leaving the question of how to achieve this budget balance in the sphere of democracy, would therefore be democratically legitimate.

---

[11] The proposal to delegate the determination of the budget balance to an independent institution has been made, for example, by Calmfors (2002, 2003) and Wyplosz (2002, 2005, 2008). The first proposal for the delegation of certain aspects of fiscal policy to an independent institution was made by Ball (1996). Independently, Gruen (1997) made a proposal very similar to that of Ball (1996).

[12] In order to avoid harmful delay in the implementation of changes in the budget balance determined by the National Fiscal Policy Committee, there should be a certain time limit for the decision making process in. If the parliament (maybe due to distributional conflicts between interest groups) has not passed a law specifying adequate changes in public spending and taxation by the end of the time limit, the committee should have the right to take action itself in order to achieve the desired change in the budget balance. One could, for example, endow the committee with the power to order a proportional change in income taxes in such a case.

That the determination of the budget balance does indeed not involve non-negligible redistributive effects is derived in the following.

Regarding potential redistributive effects associated with the determination of the budget balance, one has to distinguish between intratemporal (within a time period) and intertemporal (over many time periods) effects.

Variations in the budget balance barely have intratemporal distributional consequences. However, delegating the power to determine the budget balance to an independent institution also means delegating public debt policy (i.e. the determination of the path of public debt over time). Through public debt policy again, the fiscal authorities can shift deadweight loss associated with taxation to later generations.

However, being able to shift deadweight loss associated with taxation between generations does not mean that fiscal authorities are also able to shift welfare between generations.

If an individual thinks that the fiscal authorities do not accumulate enough public debt (i.e. that the fiscal authorities do not shift enough deadweight loss associated with taxation to later generations), they are free to increase their private consumption (i.e. to reduce the value of the inheritance to their children). If, on the other hand, an individual believes that the fiscal authorities accumulate too much public debt, they are equally free to ramp up saving (i.e. to increase the value of the inheritance to their children).

No matter what the fiscal authorities (be it the parliament or some independent institution) decide regarding the path of public debt, every individual can offset the consequences of these decisions on the distribution of welfare between himself (herself) and his (her) children by adjusting consumption. The children again are free to decide on the distribution of welfare between themselves and their own children, and so on.

That is, the question of how to distribute welfare between generations is decided upon in the best way possible: by proportional representation, i.e. everybody decides for themselves and the majority is not able to force its will upon the minority.

In order to illustrate this, assume the following change in expectations: future (in the sense of next generation) economic growth is predicted to be far higher than previously suggested. Suppose further that fiscal authorities do not change public debt policy. Given the expected increase in the welfare of their children vis-à-vis their own welfare, some individuals may decide that the size of the inheritance they were planning to leave for their children is too generous. Those individuals who think so will increase consumption (reduce saving), thereby reducing the value of the inheritance they leave their children.

In the aggregate, more consumption and less saving means that, ceteris paribus, the capital stock (i.e. the production capacities) inherited by the next generation will be smaller.

Proportional representation means that even if, at a certain point in time, the majority of people want to redistribute welfare from their children to themselves (or from themselves to their children), the minority that does not want to do so, is free to not change their consumption behaviour.

The net effect in the aggregate (i.e. the net effect on the size of the capital stock) depends on the number of people that want to increase, reduce or keep unchanged their consumption, respectively.

Because individuals can offset any effect of the fiscal authorities' public debt policy on the distribution of welfare between the current and future generations by adjusting consumption, public debt policy is in fact not able to change the distribution of welfare between generations against the will of the people. This means that the determination of the budget balance is not burdened by value judgments related to the issue of intergenerational justice.

Therefore, endowing an independent institution with the power to determine the budget balance, while leaving the question of how to achieve this budget balance in the sphere of democracy, would be democratically legitimate.

To summarize, this section proposes a new monetary and fiscal framework for the European Monetary Union. The monetary policy regime proposed is the same as in the general case: "Krugmanite" NGDPRT, as described in subsection 2.1. However, it is argued that, with respect to fiscal policy within the eurozone, a deviation from the first-best solution (as described in subsection 2.2) would be beneficial.

Each member country of the EMU should be made responsible for using fiscal policy in such a way as to ensure that, over time, the national price levels increase by the same proportion as the price level in the eurozone as a whole. As long as the sanctions on a government not meeting its mandate are sufficiently costly, the framework should provide the member countries of the eurozone with the right incentives to meet their price level targeting mandates. Alternatively, the determination of the budget balance could be delegated to an independent institution.

4 Summary

This paper proposes a new monetary and fiscal framework for the European Monetary Union. The monetary policy regime proposed for the eurozone is the same as in the general case: "Krugmanite" NGDP Rate Targeting (NGDPRT).

Under "Krugmanite" NGDPRT the central bank has the mandate to target some growth rate of NGDP. The central bank attempts to steer the actual growth rate of NGDP towards the target through the use of interest rate changes and other monetary tools. If these tools are not enough to meet the target rate of NGDP growth, the central bank commits to higher future NGDP growth in order to achieve the NGDP growth target today.

In contrast to other monetary policy regimes such as inflation targeting or even NGDP Level Targeting, "Krugmanite" NGDPRT fully offsets every aggregate demand shock – independent of its size.

With respect to the fiscal framework, the first-best solution (and the one to be implemented in the general case) would be a fiscal rule requiring the structural budget balance to be zero at all time. However, it is argued that, with respect to fiscal policy within the eurozone, a deviation from the first-best solution for a fiscal framework would be beneficial.

The basic requirement for the workability of a currency union with heavily constrained labour mobility between its member states (such as the EMU) is that the price levels in all member countries rise by the same proportion over time.

Substantial differences in the price levels of the regions of a currency area are, of course, always undesirable. The question is only whether it would be beneficial to accept other distortions in order to avoid the possibility of the emergence of substantial differences in regional price levels. The answer to this question depends, of course, on how costly the elimination of large regional price levels would be, should it be necessary.

The ongoing eurozone crisis shows that the economic and social costs associated with realigning the regional price levels in the EMU are tremendous. There is no good reason to assume that the costs associated with the removal of significant regional differences in competitiveness would be significantly lower in the foreseeable future.

This paper argues that it would therefore be desirable to accept the (limited) distortions/costs associated with deviating from the first-best solution for a fiscal framework in order to avoid the potentially huge costs associated with the possible emergence of substantial differences in regional price levels within the eurozone.

Each member country of the EMU should be subject to a fiscal rule requiring the budget balance to be set in such a way as to ensure that, over time, the national price level increases by the same proportion as the price level in the eurozone as a whole. That is, on the national level there should be a price level targeting mandate for fiscal policy. Such a fiscal framework would prevent the emergence of substantial differences in regional price levels within the eurozone in the future.

References

Alesina, A. & Tabellini, G. (2005) Why is Fiscal Policy often Procyclical?, *CESifo Working Paper Series 1556*, CESifo Group Munich.

Ball, L. (1996) A proposal for the next macroeconomic reform, *Victoria Economic Commentaries*, 14, pp. 1-7.

Barro, R. J. (1979) On the Determination of the Public Debt, *The Journal of Political Economy*, 87, 5, pp. 940-971.

Calmfors, L. (2002) *Fiscal Policy as a Stabilisation Policy Tool in the EMU*, http://people.su.se/~calmf/Fiscal_Policy_Copenhagen_May_02.pdf.

Calmfors, L. (2003) Fiscal Policy to Stabilise the Domestic Economy in the EMU: What Can We Learn from Monetary Policy?, *CESifo Economic Studies*, 49, 3, pp. 319-353.

Eggertsson, G. & Woodford, M. (2003) The Zero Bound on Interest Rates and Optimal Monetary Policy, *Brookings Papers on Economic Activity*, 34, 1, pp. 212-219.

Eslava, M. (2010) The Political Economy of Fiscal Deficits: A Survey, *Journal of Economic Surveys*, pp. 1-29.

Gruen, N. (1997) Making Fiscal Policy Flexibly Independent of Government, *Agenda*, 4, 3, pp. 297-307.
Jung, T., Teranishi, Y. & Watanabe, T. (2005) Optimal Monetary Policy at the Zero-Interest-Rate Bound, *Journal of Money, Credit and Banking*, 37, 5, pp. 813-836.

Kenen, P. (1969) The Theory of Optimum Currency Areas: An Eclectic View, in: R. Mundell & A. Swoboda (eds.), *Monetary Problems of the International Economy*, pp. 41-60 (Chicago: The University of Chicago Press).

Kouparitsas, M. A. (2001) Is the United States an optimum currency area? an empirical analysis of regional business cycles, *Working Paper Series WP-01-22*, Federal Reserve Bank of Chicago.

Krugman, P. (1998) It's Baaack: Japan's Slump and the Return of the Liquidity Trap, *Brookings Papers on Economic Activity*, 29, 2, pp. 137-187.

Lipsey, R. G. & Lancaster, K. (1956) The General Theory of Second Best, *The Review of Economic Studies*, 24, 1, pp. 11-32.

Mankiw, N. G. and Weinzierl, M. (2011) An Exploration of Optimal Stabilization Policy, *Brookings Papers on Economic Activity*, 42, 1, pp. 209-272.

Mundell, R. (1961) A Theory of Optimum Currency Areas, *American Economic Review*, 51, 4, pp. 657-665.

Ng, Y. (1983) *Welfare Economics* (London: Macmillan).

Sumner, S. B. (2011a) The case for NGDP Targeting. *The Adam Smith Institute*, 20, http://www.adamsmith.org/sites/default/files/resources/ASI_NGDP_WEB.pdf, accessed 30 June 2013.

Sumner, S. B. (2011b) Re-Targeting the Fed. *National Affairs*, 9 (Fall 2011), http://www.nationalaffairs.com/publications/detail/re-targeting-the-fed, accessed 30 June 2013.

Wyplosz, C. (2002) *Fiscal Policy: Rules or Institutions?*, http://ec.europa.eu/dgs/policy_advisers/archives/experts_groups/docs/wyplosz.pdf.

Wyplosz, C. (2005) Fiscal Policy: Institutions versus Rules, *National Institute Economic Review*, 191, pp. 64-78.

Wyplosz, C. (2008) Fiscal policy councils: Unlovable or just unloved? *Swedish Economic Policy Review*, 15, pp. 173-192.

Author contact: thomas.oechsle@cantab.net

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

The experience of three crises: the Argentine default, American subprime meltdown and European debt mess

Víctor A. Beker[1]   [University of Belgrano and University of Buenos Aires, Argentina]

Abstract
This paper aims to find out why vast masses of individuals and institutions risk their money in ventures that turn out to be a complete fiasco and to explore how to prevent this from happening again in the future. In the three cases analyzed – Argentina's 2001 crisis, the US subprime crisis and the Euro debt crisis – a common feature was the huge misjudgments by investors of the risks really involved. In at least two of these three cases, this misjudgment was induced by important actors in the financial world. In the case of Argentina, by the IMF backing of the Convertibility program; in the case of subprime mortgages, by the rating agencies' ratings.

In the case of the euro-zone, there was a general assumption that the common currency automatically meant an almost common level of risk. However, as the paper shows, in a monetary union the probability of a government default is *higher*, not smaller than for an isolated individual country government.

The fact that financial institutions have a perverse incentive to take excessive risks is emphasized. Financial activity as a whole is a public good: systemic risks to financial institutions are risks for the economy as a whole. Thus special attention should be placed on those risks capable of damaging the financial system as a whole.

The paper is divided into four chapters. The first one is devoted to the 2001 Argentine crisis; the second one, to the 2008 US financial crisis; the third one deals with the European debt crisis; the fourth one concludes.

Introduction

"Never waste the opportunities offered by a good crisis." Niccolo Machiavelli

The American financial crisis, followed by the European debt crisis, has led to increasing interest on the subject, to which economists had paid almost no attention during the optimistic years of the so-called Great Moderation.[2] However, as Reinhart and Rogoff (2009) exhaustively show, financial crises and sovereign debt defaults are far from strange events in economic history, in both less developed as well as developed countries.

While in 2003, Desai could still assert that there is a big difference in the debt management between developed and emerging countries, events after 2007 show that this is no longer valid. In spite of being endowed with a sophisticated network of financial institutions and supervisory regulatory agencies, the US economy was hit by a financial crisis that has much in common with previous emerging countries' financial crises. The same has now happened in the European Union. Moreover, the policies being undertaken by crisis countries are similar to those Argentina tried in 2001 in its desperate effort to save the peso–dollar link.

---

[1] I am grateful to James Galbraith, Bernardo Kliksberg and Edward Fullbrook for helpful comments on an earlier draft. The usual caveats apply.
[2] Notable exceptions are Bordo et al. (2001), Bordo and Murshid (2001), Eichengreen and Lindert (1989), Feldstein (1991), Reinhart and Rogoff (2003) and Sturzenegger and Zettlemeyer (2006).

Reinhart and Rogoff (2008a) demonstrate that the antecedents and aftermath of banking crises in rich countries and emerging markets have a surprising amount in common. In another paper, these authors conclude that "serial default on external debt – that is, repeated sovereign default – is the norm throughout every region in the world, even including Asia and Europe" (Reinhart and Rogoff, 2008b, 5).

As an economist living in Argentina, I was a first-hand witness of the 2001 crisis in this country. At that time, I wondered why many people in the rest of the world had thought it could not happen and had trusted their savings to the Argentine government. The impact of the 1994/5 Tequila crisis revealed the inconsistencies in the Convertibility plan and the 2001 crisis was, paraphrasing García Márquez, just a chronicle of a death foretold (see chapter I).
The same question I ask now with reference to the American and European crises. The challenge is to find out why more or less sophisticated investors risk their money in ventures that turn out to be a complete fiasco and to explore how to prevent this from happening again in the future.

Lenders do not seem to be concerned or to have the tools to properly evaluate the risks involved in their lending operations. This shows a misalignment of incentives with the public interest on both sides of the counter. On one side, lenders seem to be incentivized to take excessive risks; on the other, borrowers are driven to overborrow.

The reason why governments tend to overborrow is relatively straightforward. A government's objective function is to maximize votes. Votes are positively correlated with expenditure – it always benefits some constituency – and negatively correlated with taxes. Debt is one way of transferring payments to future governments. Therefore, governments have every incentive to maximize debt subject to the restrictions that the market imposes on them. In the real world, governments are clearly "debt biased," as Alesina and Tabellini (1990) pointed out.

In the case of private agents, the tendency to overborrow has been modeled by Bianchi (2011), who shows how optimal borrowing decisions at the individual level can lead to overborrowing at the social level. Agents fail to internalize the general equilibrium effects of their borrowing decisions on prices. This is a pecuniary externality that arises due to the presence of financial frictions.

This paper focuses on the lenders' side. I conclude that in each of the three cases analyzed here some sort of veil obscured the real risks involved. Thus, the key issue is to remove these veils and make financial markets much more transparent and accountable. Financial activity as a whole is a public good: systemic risks to financial institutions are risks for the economy as a whole. However, financial institutions per se have a perverse incentive to take excessive risks; the most aggressive institutions place pressure on the rest, and just as bad money drives out good, bad financial institutions could drive out good ones. Financial regulation should place attention on those risks capable of damaging the financial system as a whole.
In the case of public debt, a key issue is transparency in public accounts. In this respect, an independent review agency responsible for conducting performance audits and studies of selected fiscal issues may be a useful instrument for ensuring that transparency.

The present paper is divided into four chapters. The first one is devoted to the 2001 Argentine crisis; the second one, to the 2008 US financial crisis; the third one deals with the European debt crisis; the fourth one concludes.

1. Argentina's debt crisis

At the end of 1975, Argentina's external debt was 4 billion dollars; at the end of 1982, 40 billion and at the end of 2001, 140 billion. On December 2001, Argentina announced the default of its external debt. It was one of the largest defaults in present value terms since the Russian repudiation of 1918.

1.1 Introduction

In August 1982, Mexico declared that it would no longer be able to service its debt. In the wake of Mexico's default, most commercial banks reduced significantly or halted new lending to Latin America. As much of Latin America's loans were short-term, a crisis ensued when their refinancing was refused. Thereafter, Mexico, Brazil and Argentina followed suit. Argentina's default lasted until 1992, when it reached an agreement with the creditor banks within the framework of the Brady Plan.

With this antecedent, nobody thought, in the early 1990s, that Argentina's public sector could easily recover access to capital markets. However, while at the end of 1991 Argentina's gross external debt amounted to $61 billion, by the end of 1999 it had soared to $145 billion, of which 85 billion was public sector debt (Lischinsky, 2003, Table 6). How was it possible that such a serial defaulter[3] could more than double its external debt in such a short time?

This seems to contradict the explanation given by Reinhart and Rogoff (2004, 13) related to the "paradox" of why so little capital flows to poor countries; they argue that countries that do not repay their debts have a relatively difficult time borrowing from the rest of the world. However, this does not seem to have been the case for Argentina, whose external debt largely increased in the 1990s despite just coming out from default.

Of course, lenders could argue that they lent money to a country that was at that time blessed by the IMF. The IMF's point of view was clearly reflected in its former First Deputy Managing Director Anne Krueger's words during the conference on "The Argentina Crisis" in July 2002: "Between 1990 and 1997 its economy outperformed that of most other countries in Latin America, growing by more than 6 percent a year. Contagion from the tequila crisis in 1995 was severe, but short-lived with growth soon resuming. Argentina's performance was recognized internationally with President Menem's appearance alongside President Clinton at the 1999 annual meetings of the Fund and Bank." Of course, as Mrs. Krueger recognized on that occasion, there was mounting fiscal deficit but it was not then considered to be a problem; on the contrary, it was seen as an opportunity to lend money to the Argentine state, the same state that had been in default during most of the previous 10 years.

Figure 1.1 reflects the evolution of the national government public debt – external and domestic – from 1991 until 2001, when Argentina defaulted again. By that time, Argentina's total foreign currency debt was around five times the size of its annual foreign currency receipts from the exports of goods and services. Mrs. Krueger prefers to play down the role of the IMF and to put the blame on investors and lenders for "providing an apparent vote of confidence by pouring capital in." The fact is that in the 1990s Argentina was the best pupil of the IMF, the World Bank and the US government, as not only Mrs. Krueger but also the then

---

[3] Although Argentina is known as a serial defaulter, its record is surpassed by many countries in the New World and by almost as many in the Old World including France and Germany (Rogoff and Reinhart, 2004, 3).

Minister of Economy Domingo F. Cavallo recalls (Cavallo, 2004, 1). For international organizations, it was a "star" country that had followed most of the policies recommended by the so-called "Washington Consensus".

Figure 1.1 Public Indebtedness – National Government – 1991/2001 (Billions of dollars at the end of each year)
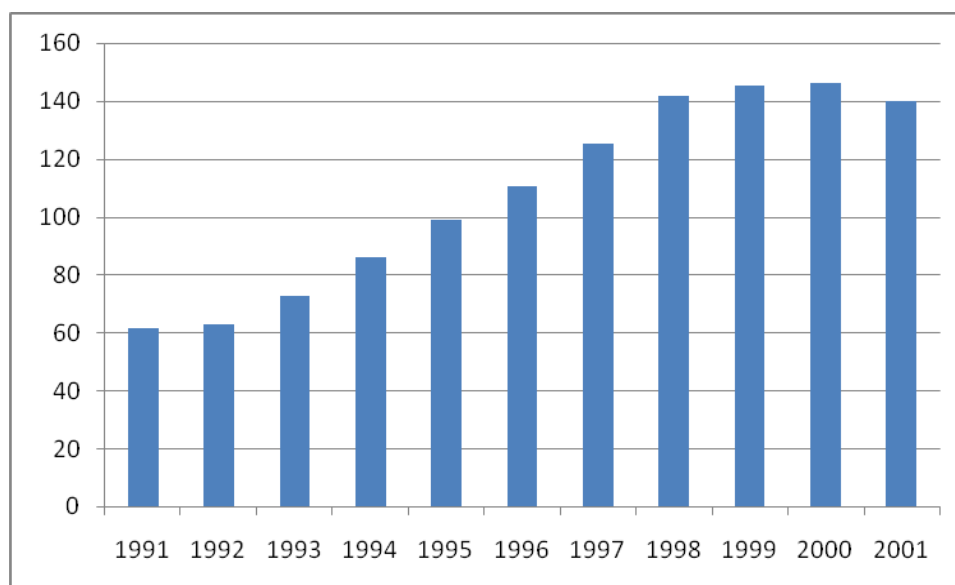
Source: Ministry of Economy

The IMF played a key role in restoring confidence in Argentina by capital markets. During the decade preceding the 2001 crisis, there were successive IMF financing arrangements for Argentina;[4] the IMF also provided extensive technical assistance during that period, dispatching some 50 missions between 1991 and 2002, mainly in the fiscal, monetary and banking areas. It widely praised Argentina for its achievements in stabilization, economic growth and market-oriented reforms under IMF-supported programs.

The misjudgment by the IMF of the sustainability of the Convertibility regime played a key role in reopening Argentina's access to capital markets. The successive bond issues that took place during the 1990s were carried out by a government whose economic policies were under the close scrutiny of the IMF, who had strongly praised them. Without its seal of approval to Argentina's economic policies, would investors and lenders have rushed to buy them? Therefore, in the case of Argentina, it seems to be clear that a primary responsibility in its 2001 public sector debt crisis was played by the IMF endorsement of an economic scheme that had been doomed to fail at least since 1995.[5]

The rest of the chapter is organized as follows. Section 1.2 is devoted to an analysis of Argentina's economic performance in the 1990s. Section 1.3 reviews the criteria for assessing the country's solvency and applies them to assess the Argentine case. Section 1.4 explains

---

[4] During the 1990s, there were four IMF arrangements: arrangement under the Extended Fund Facility (EFF) approved on 3/31/92; Stand-By Arrangement (SBA) approved on 4/12/96; arrangement under the EFF, approved on 2/4/98; and SBA, approved on 3/10/00.
[5] See Beker and Escudé (2008, 23/24).

the reasons for Argentina's growing public sector debt. Section 1.5 is devoted to analyzing the role of the IMF in the Argentine debt crisis. Section 1.6 summarizes.

1.2 Argentina's economic performance in the 1990s

The economic performance of Latin American countries in the 1980s was unsatisfactory. In what has been called the "lost decade," the region's economy was disrupted by the debt crisis and raging inflation. This experience shocked the region; as a result, Latin America embraced structural economic reforms during the 1990s. All countries liberalized international trade and external capital flows and privatized public utilities. Argentina was no exception. Actually, it was one of the countries where more aggressive economic reforms were implemented.

After the hyper-inflationary processes of 1989 and 1990, drastic economic reforms took place in Argentina. The key measures that shaped this economic program were the Convertibility Law, the liberalization of external trade and financial flows and the privatization of public enterprises.

The Convertibility Law established a fixed exchange rate of one peso to one dollar. The Central Bank was obliged to sell foreign currency at that rate as required by the market. In order to fulfill this obligation, it had to keep international reserves equivalent to at least 100% of the monetary base (except for up to 10% of the monetary base which could be backed by dollar-denominated government bonds). This meant, virtually, the transformation of the Central Bank into a Currency Board. As a result of this package, inflation was drastically abated from a level of 5% per year in 1989 to just 0.16% in 1996. Moreover, GDP grew by 40% between 1990 and 1994.

Trade liberalization was reflected in a huge increase in foreign trade. Imports soared, from $4.1 billion to $21.6 billion in 1994, while exports rose from $3.7 billion to $20.1 billion in the same period. The participation of imports in aggregate supply expanded from 5.6% in 1990 to 14.6% in 1994. This increase in international trade was accompanied by substantial expansion in the deficit in the current account.

Convertibility together with trade liberalization assured the stability of tradable goods prices. Their domestic prices would not increase if international prices did not because imports could easily replace local production. However, this did not mean the stability of prices for non-tradable goods. The result was a change in relative prices in favor of non-tradables.

Excess demand – fueled by foreign capital inflow – resulted, on one hand, in an increase in the volume of imports and, on the other hand, in the price of non-tradable goods. This price behavior resulted in the continuous erosion of the competitiveness of tradable sectors. The current account deficit thus increased from $5.4 billion in 1992 to $10.1 billion in 1997 – more than one-third of that year's exports.

Therefore, the current account deficit kept growing during the 1990s and more and more capital inflows were needed to make up for it. As can be seen in Table I.1, Argentina needed a surplus of around $10 billion per year in its capital account not to lose foreign exchange reserves. When there was a sharp reduction in global capital flows to emerging market economies, as happened in 1995 due to the Tequila effect, real GDP fell 4.6% and unemployment soared, reaching 17%.

Contrary to the conclusion by the IMF and mainstream analysts that Argentina's economic behavior in the presence of the Tequila effect proved to be the strength of its economy, it showed its Achilles' heel: its high sensitivity to external flows. As later events proved, 1995 was a general rehearsal for the 2001 crisis. The lack of access to funds on international capital markets would strangle the Argentine economy under the Convertibility regime. In the late 1990s, the Argentine economy suffered a series of external shocks: the East Asian crisis in 1997, the Russian one in 1998, the 1999 devaluation of the Brazilian real, which had a negative impact on the competitiveness of Argentina's significant exports to this country, and the appreciation of the US dollar against most other currencies, which increased Argentina's real effective exchange rate. In the presence of all these adverse shocks, the Convertibility regime prevented a flexible domestic policy response.

The twin deficits required continuous access to external financing. However, far from being considered a drawback or a weakness of the economic program, they were considered to be just a minor detail, assuming that foreign capital markets would always be available to finance both disequilibria. An increasing stock of external debt, rising country risk premiums and sluggish growth caused the ratio of debt to GDP to rise uncontrollably until the default came in 2001.

Sovereign debt has usually been assumed to be almost risk-free because it is supposed that governments can always resort to an increase in taxes to service it. However, in the real world there is always a political limit for that.

Table 1.1 Balance of Payments – Argentina 1992/97

	1992	1993	1994	1995	1996	1997
Current Account	-5.462,0	-7.672,0	-10.117,0	-2.768,0	-3.787,0	-10.118,8
Commodities	-1.450,0	-2.426,0	-4.238,0	2.238,0	1.622,0	-3.194,6
Exports	12.235,0	13.117,0	15.839,0	20.964,0	23.811,0	25.223,0
Imports	13.685,0	15.543,0	20.077,0	18.726,0	22.189,0	28.417,6
Services	-2.257,0	-2.730,0	-2.941,0	-2.222,0	-2.495,0	-3.069,2
Exports	2.454,0	2.454,0	2.599,0	2.860,0	3.226,0	3.271,0
Imports	4.711,0	5.184,0	5.540,0	5.082,0	5.721,0	6.340,2
Income	-2.416,0	-2.927,0	-3.258,0	-3.216,0	-3.248,0	-4.205,0
Interest	-1.289,0	-1.081,0	-1.136,0	-1.054,0	-1.326,0	-1.770,0
Earned	2.099,0	2.135,0	3.073,0	4.348,0	4.587,0	5.332,0
Paid	3.388,0	3.216,0	4.209,0	5.402,0	5.913,0	7.102,0
Profits & Dividends	-1.127,0	-1.846,0	-2.122,0	-2.162,0	-1.922,0	-2.435,0
Transfers	661,0	411,0	320,0	432,0	334,0	350,0
Capital Account	8.567,3	12.152,0	10.678,0	2.699,0	7.569,0	13.180,8
I. Banking Sector	826,0	-1.528,0	1.895,0	4.360,0	-519,0	-1.794,0
Central Bank	-177,0	-2.818,0	307,0	1.929,0	849,0	-800,0
Others	1.003,0	1.290,0	1.588,0	2.431,0	-1.368,0	-994,0
II. Public Sector	1.285,0	7.121,0	4.097,0	5.945,0	8.731,0	7.331,0
National Govt	1.853,0	6.473,0	4.471,0	6.435,0	8.583,0	6.495,0
Local Govt	31,0	889,0	189,0	374,0	612,0	1.231,0
Public Enterprises	-599,0	-241,0	-563,0	-864,0	-464,0	-395,0
III. Private Sector	2.766,0	4.559,0	4.454,0	4.923,0	5.415,0	9.035,0
IV. Other Movts	3.690,3	2.000,0	232,0	-12.529,0	-6.058,0	-1.391,2
Variation in Reserves	3.105,3	4.480,0	561,0	-69,0	3.782,0	3.062,0
MEMO ITEM						
Imports	14.873,0	16.783,0	21.590,0	20.122,0	23.761,0	30.323,6

Source: Ministry of Economy

1.3 Country's solvency and the Argentine case

Although no simple rule can help determine when foreign debt accumulation is sustainable or not, a number of criteria can be used in assessing the sustainability of the foreign debt of a country. The issue is summarized in Roubini (2001, 3–4).

The analytical literature on current account and foreign debt sustainability provides a theoretical criterion that is not particularly stringent. As long as the discounted value of trade balances is at least equal to its initial foreign debt, the country is solvent; this means only that the country cannot increase its foreign debt faster than the real interest rate on this debt. Therefore, any path of the current account such that the infinite sum of all current accounts is equal to the initial foreign debt of the country is consistent with solvency. This means, for instance, that if the real interest rate is greater than the rate of the growth of an economy, solvency is consistent even with a foreign debt to GDP ratio that grows continuously over time.

A similar criterion applies in determining whether the public debt of a government is sustainable or not. Specifically, as long as the discounted value of primary balances is at least equal to the initial public debt, the public sector is solvent. However, the dynamics of the current account that lead to an increase without bounds of the foreign debt to GDP ratio can be seen as effectively unsustainable: financial markets will eventually become concerned about the country's ability and willingness to repay its debt and will limit its borrowing, leading to a foreign debt crisis. The same things apply for the case of domestic debt.

That is why a non-increasing foreign debt to GDP ratio has been seen as a practical sufficient condition for sustainability: a country is likely to remain solvent as long as the ratio is not growing. Similarly, public debt can be viewed as sustainable as long as the public debt to GDP ratio is non-increasing. The "resource balance gap" is thus the difference between the current trade balance and the trade surplus required to stabilize the debt to GDP ratio. In the same way, the fiscal "primary gap" is the difference between the fiscal primary balance and the primary balance required to stabilize the debt to GDP ratio. This criterion provides a normative rule: how much a trade surplus or primary surplus is required to close the resource or primary gap. However, it does not directly provide a tool to assess whether a certain stock of debt is sustainable or not.

Several alternative indicators of fiscal and external debt sustainability can be used to assess insolvency. Three of the most commonly used are the debt to GDP ratio, the debt to export ratio and the debt to government revenue ratio. The relevant denominator depends on the constraints that are most binding in an individual country, with GDP capturing overall resource constraints, exports those on foreign exchange and revenues those on the government's ability to generate fiscal resources. In general, it is useful to monitor external debt in relation to GDP and export earnings and public debt in relation to GDP and fiscal revenues.

In this respect, the analysis by IMF staff for low-income countries yields a threshold value for the debt to GDP ratio of around 43%, 192% for the debt to exports ratio and 288% for the debt-to-revenue ratio (IMF and IDA, 2004, 57).

Based on the criterion of the external debt to GDP ratio, Argentina crossed the threshold in 1998 (Table 1.2). However, the GDP calculation was biased upwards by the overvaluation of the peso, so entrance into the "danger area" might have happened a couple of years before.

Concerning the debt to export ratio, Argentina had in 2001 a ratio of 561%, well above the threshold value, although the same happened with all the values of this series in the 1990s (see Table 1.3). Finally, the debt to government revenue ratio was 220% in 2001, below the threshold value for this coefficient.

Therefore, the coefficients themselves do not explain why Argentina defaulted in 2001. If the relevant coefficient were the debt to exports ratio, Argentina was already a potential defaulter in 1991. However, it managed to borrow almost $80 billion during the following 10 years, more than doubling its external debt. Perhaps the most important issue at the time of default was the high share of short-term external debt. In fact, for both 2002 and 2003, the repayment of principal exceeded 80% of the exports. Adding interest payments of about $12 billion, total debt servicing largely exceeded annual exports. Argentina depended on creditors' willingness to roll over its external debt. This became increasingly difficult since capital flows to Argentina quickly decelerated after the 1998 Russian crisis. By mid-2001, the economic authorities initiated a process to improve the maturities by extending them. A $30 billion government debt swap took place in June. The government thought this transaction would offer financial relief in terms of the repayment of principal and interest payments of around $4.5 billion annually. However, this was carried out at the price of accepting an implicit interest rate of 15%,[6] which was interpreted by creditors as announcing a high probability of default. After that, the failure of a Treasury bill auction confirmed that the Argentine government had lost access to credit. Default was then inevitable.

Argentina had been continuously issuing new bonds to cancel most of the principal and interests of the debt that were becoming due. Only when default was imminent did creditors refuse to go on playing this game. Even then – in September 2001 – the IMF approved one last significant tranche of financing for Argentina.

In the analysis of its role in the Argentine crisis, the IMF (2003, 72) poses the dilemma its authorities faced at that time: even after realizing the high probabilities of failure, it went on supporting the Argentine economic program in light of the high and immediate costs of withdrawing support. This reflects the path dependency existing in decision making: once you make a considerable wrong bet, you are doomed to increase it in order to try to save your initial investment.

In the context of political instability – the governing coalition was undergoing a political crisis since the resignation by the vice-president in October 2000 – Argentina finally defaulted at the end of 2001 after the then president resigned from his job.

1.4 The reasons for Argentina's growing public sector debt

When the Convertibility plan started in 1991, a restriction was placed on the Central Bank. It could not make loans to the government (except for short-term limited amounts). Given the reluctance by foreign lenders at that time to become involved in Argentina, it was taken for granted that that constraint was practically equivalent to excluding the possibility of running a fiscal deficit. During the transition, the government would resort to the proceedings from privatizations while leveling expenses with revenues. In fact, in 1993 – for the first time in decades – the nonfinancial public sector had no deficit. However, exactly at that time it was decided to reform the social security system. The main effect of this reform was to transfer most of the system revenues to the private sector, while keeping most of the expenses within

---

[6] At that time, the one-year US Treasury interest rate fluctuated around 3.6%.

the public sector. This meant that since 1994 the federal budget was again continuously in deficit, even in years of good economic growth. New debt was added to old debt year after year, and debt plus interest grew much faster than the economy. At the end of 1994, the federal government's gross debt was $75 billion, while GDP in 1994 was $257 billion. By the end of 2001, debt was almost twice as large, $140 billion, while GDP was only $271 billion, just 5% higher than in 1994.

Thus, what in 1991 was unthinkable did happen: since 1994, Argentina had recovered access to international capital markets. Therefore, the constraint placed on the Central Bank became non-binding. Capital markets were willingly available to finance Argentina's public sector debt. How did this Copernican change happen? First, since 1992 Argentina was under the umbrella of an IMF-supported program; second, it enthusiastically adhered to the Washington Consensus and its principles; third, the Currency Board was a guarantee of no devaluation; finally, high interest rates were a significant attraction. From 1994 on, what has been called a "bond festival" took place until the 2001 default put an abrupt end to it.[7]

Table 1.2 – External debt to GDP ratio   **Table I.3 –** External debt to export ratio

Year	%
1991	32,35
1992	27,36
1993	30,53
1994	33,28
1995	38,20
1996	40,32
1997	42,61
1998	47,11
1999	51,54
2000	51,91
2001	51,95

Source: Ministry of Economy and IMF

Year	%
1991	503,54
1992	512,06
1993	511,53
1994	550,47
1995	540,78
1996	470,10
1997	460,95
1998	472,30
1999	532,97
2000	627,32
2001	560,60

Source: Ministry of Economy and IMF

The continuous support by the IMF to the Argentine program, even after the Tequila crisis showed its high sensitivity to external flows, allowed the government to pile up a huge debt, long after it was evident that the Currency Board was unsustainable. Therefore, it is worthwhile analyzing the role of the IMF in the Argentine crisis.

1.5 The role of the IMF

After the 2001 crisis, the IMF produced two documents. One was aimed at examining the origins of the Argentine crisis and its evolution until early 2002 (IMF, 2003). The second one was produced by the IMF's Independent Evaluation Office (IEO); this evaluated the role of the IMF in Argentina during 1991–2001 (IMF, 2004).

---

[7] Those interested in Argentina´s development after the 2001 devaluation and default may have a look at Frenkel (2012).

In its report, the IEO recognizes that "the catastrophic collapse of the Argentine economy in 2001–02 represents the failure of Argentine policymakers to take necessary corrective measures at a sufficiently early stage. The IMF on its part, supported by its major shareholders, also erred in failing to call an earlier halt to support for a strategy that, as implemented, was not sustainable" (IMF 2004, 64). The IEO argues that favorable factors allowed the exchange rate regime to survive for a number of years without being severely tested but the situation changed in 1998–99 when Argentina was hit by a series of adverse shocks. However, it admits that "these shocks would have been difficult enough to handle at any time, given the rigidity of the fixed exchange rate and the lack of downward flexibility in domestic wages and prices" (Ibid.). Therefore, it recognizes that the Convertibility regime, because of its rigidity, was incapable of dealing with any adverse shock of a certain volume. In fact, it did not pass the Tequila test: in order to keep a fixed exchange rate, the country's economy experienced a jump in its unemployment rate from 10.7% to 18.4% between May 1994 and May 1995 (see Table 1.4). Therefore, if a test was needed, the Mexican crisis provided it. However, the IMF interpretation *even in 2003* was that "the economy had successfully weathered the Tequila crisis of the mid-1990s" IMF (2003, 3). Calling a success the management of a crisis that meant an 80% increase in the unemployment rate is evidence that the IMF underweights unemployment in its assessment scheme.

The IEO report goes onto admit that the IMF's "support gave credibility to Argentina's stabilization and structural reform efforts" (IMF, 2004, 65), although the IMF was initially skeptical as to whether the Convertibility plan would work. This suggests that political considerations prevailed over the technical opinions of IMF staff.[8]

Table I.4 Unemployment rates – May and October 1991/95

Year	May	Oct
1991	6,9	6,0
1992	6,9	7,0
1993	9,9	9,3
1994	10,7	12,2
1995	18,4	16,6

Source: INDEC

Although the report underlines that the IMF correctly identified the potential vulnerabilities inherent in the Convertibility regime, the fact is that in spite of that the IMF went on supporting that regime even after it was clear that this support allowed Argentine authorities to swiftly increase Argentine public debt to unsustainable levels. "Moreover, the IMF ... began to endorse the exchange rate regime itself. Indeed, the IMF publicly lauded convertibility as an example of a Currency Board, the only type of fixed exchange rate regime that is fundamentally sustainable in a world of high capital mobility" (IMF, 2004, 65). In summary, the IMF's support was a necessary element without which it would have been difficult to increase Argentina's indebtedness as occurred during the 1990s.

---

[8] "…dissenting views were overruled by such considerations as the need to maintain influence with a member country or a desire to preserve the catalytic effect of the IMF's seal of approval" (IMF, 2004, 66).

1.6 Summary

After the hyper-inflationary processes of 1989 and 1990, drastic economic reforms took place in Argentina. The central piece of this program was the Convertibility Law, which established a fixed exchange rate of one peso to one dollar. The Central Bank could issue domestic currency only against foreign currency and could not make loans to the government except for a very tiny sum. It was taken for granted that this constraint was practically equivalent to excluding the possibility of running a fiscal deficit. However, soon this proved not to be true: from 1994, Argentina recovered access to international capital markets and since then increased its public debt at a very fast rate. How were lenders convinced to lend huge amounts of money to a serial defaulter such as Argentina? There is no explanation but endorsement by the IMF of Argentina's economic program. It is hard to believe that lenders would have rushed to buy Argentine bonds without the IMF's seal of approval.

The misjudgment by the IMF about the sustainability of the Convertibility regime played a key role in reopening Argentina's access to capital markets. Continuous support by the IMF to the Argentine program, even after the Tequila crisis showed its economy's high sensitivity to external flows, allowed the government to pile up a huge debt, long after it was evident that the Currency Board regime was unsustainable. The IMF played in the Argentine case the same role as credit rating agencies played in the 2008 American crisis: to induce lenders to put their money into buying securities of doubtful collectability.

2. The American financial crisis

2.1 Introduction

The core of the 2007/08 financial market crisis has been the discovery that many securities were actually far riskier than people originally thought they were. The process of securitization allowed trillions of dollars of risky assets – subprime mortgages in the first place – to be transformed into securities that were widely considered to be safe.

I have defined elsewhere (Beker, 2010, 5) the American financial crisis as a typical case of professional malpractice, an extended malpractice by hundreds of professionals in banks and rating agencies who created and certified as almost risk-free securities assets that were actually highly risky as the events after 2007 overwhelmingly showed.

Subprime mortgage securitization models relied on assumptions and historical data that turned out to be incorrect and therefore made incorrect valuations. Substantial lending to subprime borrowers was a recent phenomenon and historical data on the defaults and delinquencies of this sector of the mortgage market was scarce (see Coval et al., 2009, 15). Some models were not even based on historical data because they referred to transactions for which there was no active trading market. "The mathematical rigor, elegance and the numerical precision of the various risk-management and asset-pricing tools have a tendency to 'hide' the weaknesses of these models and their underlying assumptions, which are necessary to guarantee the models' values to those who have not developed them" (Schneider and Kirchgässner, 2009, 6).

Securitization enabled mortgage lenders to sell off loans as they were made, thus creating moral hazard since this meant that lenders could pass along the risk of default to investors.

Mortgage underwriting standards fell once lenders did not have to live with the credit consequences of their loans. Gorton (2008, 2009) disagrees with this interpretation, arguing that many lenders went under after the crisis. However, one may use just the opposite argument: they went under precisely because the crisis burst out before they were able to distribute all the securitized debt they had created. On the other hand, lenders who did not sell all the loans they originated were able to buy relatively inexpensive protection against credit risks through credit default swaps, which was another way of transferring risk to a third party.

Securitization, which was supposedly aimed at reducing informational asymmetry,[9] became a tool to take advantage of that asymmetry. Because of the asymmetric information between the lender and the investor, rating agencies came on scene to provide the latter with accurate risk evaluation. However, the problem was that rating agencies are paid by the issuer not by the investor. This raised a conflict of interest, as was clearly exposed by the high credit ratings given to actually highly risky assets.

Behind this extended malpractice was the pressure caused by a liquidity glut, which forced lenders to compete aggressively for business. That is why monetarists blame exclusively the money glut for the crisis. However, the existence of a money glut is a necessary but not sufficient condition for developing a bubble like the one that culminated in the crisis. It was the combination of this money glut with financial deregulation that was lethal for the American economy.

2.2   The money glut

Global current account imbalances have been singled out as a key factor contributing to the global financial crisis. Current account surpluses in several emerging market economies (China and other Asian countries plus oil-exporting countries) are said to have helped fuel the credit booms and risk taking in the major advanced deficit countries at the core of the crisis, by putting significant downward pressure on world interest rates and/or by simply financing the booms in those countries.

Bracke and Fidora (2012) test the global liquidity glut hypothesis versus the global savings glut one. They find that US monetary policy shocks explain the largest part of the variation in imbalances and financial market prices. Savings shocks and investment shocks explain less of the variation. Hence, according to them, a "liquidity glut" may have been a more important driver of the real and financial imbalances in the US and emerging Asia that ultimately triggered the 2007–08 global financial crisis.

Borio and Disyatat (2011, 20) reject the dominant "excess saving" view, arguing that "the saving-investment framework describes the real side of the economy. The equality between ex ante saving and investment is an equilibrium condition for the goods market." For them, the focus of the analysis should be placed on monetary policy: "It is monetary policy that underpins the term structure of market interest rates" (Ibid., 24).

They argue that "the geographical breakdown of capital inflows into the US in the run-up to the crisis is hardly consistent with the excess savings view. By far the most important source of capital flows was Europe, not emerging markets. Of this, more than half came from the United Kingdom, a country running a current account deficit, and roughly one-third from the

---

[9] See, for example, Schwarcz (2011, 4).

euro-area, a region roughly in balance. This amount alone exceeded that from China and by an even larger margin that from Japan, two large surplus economies." (Ibid., 15). They refer to Milesi-Ferretti (2009), according to whom on the eve of the crisis (June 2007) the holdings of privately issued mortgage-backed securities were concentrated in advanced economies and offshore centres. Contrary to the excess savings hypothesis, they maintain that "the focus on global current account imbalances misses the role of European banks in supporting the boom in US housing credit and the subsequent collapse of such financing" (Ibid., 20). They conclude that the main contributing factor to the financial crisis was the "excess elasticity" of the international monetary and financial system. They argue that "the financial system can endogenously generate financing means, regardless of the underlying real resources backing them. In other words, the system is highly elastic. And this elasticity can also result in the volume of financing expanding in ways that are disconnected from the underlying productive capacity of the economy" (Ibid., 28).

However, it does not seem that one approach necessarily excludes the other. The "excess saving" view may explain the huge increase in the official holdings of US Treasury securities by the countries that accumulated foreign exchange reserves during the past decade. This allowed the American economy to run twin deficits during these years and keep interest rates low. These low interest rates were validated by an expansionary monetary policy (the "liquidity glut").

The real issue is not the ultimate cause of cheap money but why low interest rates did not stimulate investment in real productive capacity instead of feeding a colossal speculative bubble in the real estate market. The answer is that speculation promised greater benefits with almost no risk. Here comes the key role played by rating agencies. Trillions of dollars of risky assets were transformed into mostly AAA-rated securities. This was the key element in feeding the subprime mortgage bubble: "The three credit rating agencies were key enablers of the financial meltdown. The mortgage-related securities at the heart of the crisis could not have been marketed and sold without their seal of approval" (FCIC, 2011, XXV).

2.3   The role of credit rating agencies

Credit rating agencies were an essential input into the process of manufacturing vast quantities of triple-rated securities with attractive yields. In a period of low interest rates, they were eagerly bought up by investors unaware of the real risks they entailed.

Risks were strongly mispriced. Investors thought they had bought a Mercedes Benz; it took a certain time for them to find out they were just "lemons." Coval et al. (2009) explain thoroughly the roots of rating agencies' errors and why they were unable to accurately assess securities risks, in particular systematic risks.

While house prices kept rising, risks stayed hidden. If an owner could not meet the monthly payments, the bank renegotiated the mortgage. The renegotiation would raise the principal to the new higher house value in exchange for lowering the monthly payment. Therefore, the delinquency rate was low.

The situation changed abruptly when house prices started falling. Most borrowers who could not afford the monthly payments had no alternative but to default their subprime mortgages, as many of them found themselves holding mortgages in excess of the market values of their

homes. Subprime-related securities experienced large losses; investors learned the hard way how risky these assets were.

Few investors had been worried that the underlying assets might be overvalued. This is not surprising taking into consideration that credit rating agencies evaluated and deemed them to be "safe." Therefore, there is no mystery why investors massively rushed to buy these "toxic" assets. The originating banks, which were presumably able to charge a higher interest rate, and many US institutional investors, who needed high ratings to buy the securities at all, both had a vested interest in rating agencies awarding high ratings.

A theoretical argument advanced by Kartik et al. (2005) may help explain rating agencies' behavior. In the context of an analytical model of communication games, the authors assume a setting in which the sender of a message is interested in the *average* response of a population of receivers characterized by heterogeneous strategic sophistication. They demonstrate that in such cases there is a unique non-decreasing, differentiable separating equilibrium. This equilibrium has the important property that in every state of the world, the sender induces a belief in naive receivers such that the average population response is in fact his or her bliss point. That is, the sender can achieve his or her first-best outcome in such a setting, even though sophisticated receivers correctly infer the state of the world in equilibrium. In the equilibria they identify, the message sent by the sender has a literal meaning that is inflated, a literal meaning higher than the true state of the world. Nevertheless, a sophisticated receiver correctly infers the true state by inverting the observed message according to the equilibrium language. A credulous receiver instead interprets the equilibrium messages with some non-equilibrium-based rule and is accordingly deceived, taking biased actions. If naive receivers are on one side of the playing field and sophisticated ones are on the other one, we get something like the subprime meltdown.

2.4   The role of banks

Banks pursued an aggressive lending policy in order to get rid of the excess money in their vaults. As stated before, securitization created moral hazard since it meant that lenders could pass along the risk of default to investors or insurance companies. This encouraged excessive risk taking. The problem was magnified because the most aggressive institutions put pressure on the rest of them: safe institutions that desired to be more careful and scrutinize more deeply the repayment capacities of their potential customers would lose market share and make fewer loans. Just as bad money drives out good, bad financial institutions could drive out good ones.

The rapid increase in market share by unregulated brokers and originators put pressure on regulated banks to lower their underwriting standards. Securities backed by subprime mortgages lent to borrowers whose abilities to repay were doubtful became prominent in the banking business.

As documented in Fratianni and Marchionne (2009), large US banks were dominant in securitization. The ratio value of securitization activities – covering real estate loans, credit cards receivables, automobile loans, other consumer loans and commercial and industrial loans – of total bank assets for large US banks started at 14.5% at the end of 2002 and reached a peak of 18.6% in the first quarter of 2007. By contrast, the securitization/asset ratio for intermediate-sized banks was below 1%, while small US banks were not materially involved in securitization. A similar pattern holds for derivatives.

Securitization gave birth to a complex shadow banking system to intermediate credit through a wide range of securitization and secured funding techniques such as asset-backed commercial papers, asset-backed securities, collateralized debt obligations (CDOs) and repurchase agreements (repos). The shadow banking system – developed out of the regulated banking system – comprises securitization vehicles, asset-backed commercial paper vehicles, money market funds, investment banks, mortgage companies and a variety of other entities. It provided sources of funding for credit by converting opaque, risky, long-term assets into money-like, short-term liabilities (Pozsar et al., 2012, 1). Therefore, credit intermediaries relied on short-term liabilities to fund illiquid long-term assets. In the shadow banking system, loans, leases and mortgages were securitized and thus they became tradable instruments. Funding was also in the form of tradable instruments, such as commercial papers and repos. However, the shadow banking system was presumed to be safe due to the liquidity and credit puts provided by the private sector. These puts **underpinn**ed the perceived risk-free, highly liquid nature of most AAA-rated assets that collateralized credit repos and shadow banks' liabilities more broadly (Ibid., 2).

The shadow banking system emerged from the transformation of the largest banks from low return-on-equity institutions that originate loans and hold and fund them until maturity with deposits to high return-on-equity entities that originate loans in order to warehouse and later securitize and distribute them, or retain securitized loans through off-balance sheet asset management vehicles (Ibid., 15). This allowed banks to conduct lending with less capital than if they had retained loans on their balance sheets. This process enhanced the return on equity of banks, or more precisely, of their holding companies. Moreover, it enabled them to bypass existing regulations regarding minimum capital ratios. The funding and maturity transformation of structured credit assets was not only conducted from the US, but also from Europe and offshore financial centers.

The gross measure of shadow bank liabilities grew to nearly $22 trillion in June 2007, while traditional banking liabilities were around $14 trillion in 2007 (Ibid., 9).[10] At the beginning of the 1990s, both types of liabilities totaled practically the same.

When the housing bubble exploded in 2007, real estate markets went down together and mortgage defaults soared in Florida as well as in California. Mortgage-backed securities carried the dual risk of high rates of default due to the low credit quality of borrowers and the high level of default correlation as a result of pooling mortgages from similar geographical areas and vintages. When prices fell in the home market, subprime-related assets deteriorated. Repo depositors became concerned about the solvency of their counterparties. In the summer of 2007, panic started in the repo market, which suffered a run when depositors required increasing haircuts. In the repo market, depositors and borrowers are individually matched; each depositor gets his or her own collateral. Firms – often money market funds and corporations – deposit short-term cash; all types of securitized products are used as collateral. The haircut is the percentage difference between the market value of the pledged collateral and the amount of funds lent. The size of the haircut reflects the credit risk of the borrower and the riskiness of the pledged collateral. Depositors can "withdraw" their funds by not rolling over their repo agreements and returning the collateral, or they can withdraw by increasing the haircut on the collateral. Haircuts were zero until August 2007.

---

[10] This led Krugman (2009, 170) to call it ¨the non-bank banking crisis¨.

After that, haircuts rose and continued to rise; some asset classes became simply unacceptable in repo (Gorton, 2009, 30/33).

There was a flight to quality. It was not known which counterparties were really at risk and consequently there was a run on all banks. Defaults and losses on other loan types also increased significantly as the crisis expanded from the housing market to other parts of the economy. The rest is a well-known story.

It is clear that the roots of the problem were the subprime mortgages recklessly provided to doubtful borrowers. This behavior was stimulated by the "originate-to-distribute" model implemented through the shadow banking system. As stated above, Gorton rejects this hypothesis; in support of his argument, he exemplifies that in 2006 and early 2007 some banks kept the most senior proportions of CDOs on their balance sheets.[11] In the same vein, he argues that when loans are sold in the secondary market, the mortgage servicing rights created are typically not sold. Although he admits that underwriting standards were lowered, he contends that it seems difficult to define a decline in lending standards. Gorton (2008, 67) argues that the design of subprime mortgages and subprime securitizations are unique in that they are particularly sensitive to declines in house prices: "The key security design feature of subprime mortgages was the ability of borrowers to finance and refinance their homes based on the capital gains due to house price appreciation over short horizons and then turning this into collateral for a new mortgage" (Gorton, 2008, 3). However, when house prices began to slow their growth and ultimately fell, the value of the chain of securities began to decrease. Gorton seems to argue that banks were not recklessly selling loans to doubtful debtors but that they were confident that house prices would never significantly decline. If so, they were justified in not taking care with underwriting standards and even keeping part of the risky assets in their portfolio without taking full advantage of the risk-minimizing originate-to-distribute system.

It is difficult to identify which of these hypotheses is right. Perhaps it was a combination of both. Anyway, it is clear that for one reason or the other – or both – banks had no incentives to carefully monitor the loans they were selling. Gorton (2008, 73) includes a table showing that mortgages with less than full documentation soared from 28.5% in 2001 to 50.8% in 2006.

On top of this, a gigantic interlinked structure of securities was created with the help of the mostly unregulated shadow banking system. This structure was ready to amplify and spread to the whole financial system following the failure of any mortgage loan. The conditions were right for a perfect storm. When house prices finally stopped rising, borrowers could not refinance their way out of financial difficulty, mortgage defaults soared and the whole securitization building collapsed.

Every sector of the financial services industry was vulnerable to the effects of the toxic mortgage contagion. Then, the next question is why regulators did not foresee the likely storm or if they did why did not act to prevent it.

---

[11] In fact, after manufacturing some security, some banks used to sell the highest risk tranches and retain some of the super senior position. This only proves that they thought that only the junior tranches were highly risky.

2.5   The role of regulators

When asked how such huge mismanagement in the mortgage market could have happened, the first line of defense by regulators has been to argue that most of the problems originated in someone else's jurisdiction. In fact, financial activity regulation is deeply fragmented in the US. There are at least 10 different types of institutions in charge of regulating the activities involved in the subprime meltdown. This is just an invitation to take advantage of the gaps such a partitioned system provides.

In spite of this plethora of regulating institutions, there was no statutory regulator for investment bank holding companies and the shadow financial system was mostly unregulated. A mortgage lent by a holding company affiliate was subject to very light regulation; a mortgage lender or a broker unaffiliated with a bank was virtually unregulated. The 2000 Commodity Futures Modernization Act specifically prohibited swaps regulation. However, several subprime participants that performed poorly were in fact regulated by one of the banking agencies but the relevant banking agency failed to compel the institution to adequately comply with guidance (Robertson 2011, 17).

Obviously, with over 10 million mortgage applications for home purchases in 2006 and millions of mortgages making their way into mortgage-backed securities every year, it was not even remotely feasible to inspect every mortgage (Ibid., 20). However, in such a case, any auditor could have taken a test sample from any portfolio and, through the re-verification of several loan items, estimate the credit quality of the portfolio relative to its advertised quality.[12] Even more, precisely the huge number of mortgages lent by each institution every month should have made regulators suspect that the loans were not subject to due assessment. Unless they thought that due assessment was a waste of time because they believed that housing prices were going to rise forever.

The US Securities and Exchange Commission (SEC) was in charge of regulating or overseeing almost 35,000 financial firms and public companies. Each entity issuing an asset-backed security had to file a prospectus with the SEC, "a prospectus, which typically can be as long as 300 pages for a single security, contains an impressive amount of data regarding the asset pool. Regrettably, all of this prospectus information is unverified" (Ibid., 22).

Insurance companies such as AIG were subject to state insurance regulators. However, it seems that nobody noticed that AIG wrote $656 billion in credit insurance on structured finance products with only $54 billion in resources to pay those claims (Ibid., 35/36).

National banks and their operating subsidiaries as well as the federal thrifts and their operating subsidiaries were subject to exclusive federal supervision by the Office of the Comptroller of the Currency (OCC) and the Office of Thrift Supervision, respectively.

State-chartered banks and thrifts and non-bank affiliates of bank and thrift holding companies were subject to both federal and state supervision, while mortgage lenders not affiliated with banks or thrifts were subject only to state supervision.

However, there is no substantial difference in the results achieved by these different regulators: 22% of the non-prime loans originated by national banks and their subsidiaries subsequently entered the foreclosure process at some time after origination, while the market

---

[12] In fact, this is what Robertson (2011) proposes to do in the future.

average was 25.7% for those types of loans (Dugan, 2010, 9). The slightly lower percentage of failures in the case of national banks does not include the defaulted subprime loans made through nonbank institutions. In fact, a number of large bank holding companies owning national banks often used nonbanks for their subprime lending (Ibid., 7).

In his testimony before the Financial Crisis Inquiry Commission (FCIC), the then Comptroller of the Currency argued that "most securitizations and structured credit activities have been conducted outside of banking subsidiaries in holding company affiliates registered as broker-dealers and regulated by the SEC and the Federal Reserve" (Ibid., 13).

Although the Federal Reserve's supervisory capital assessment program, popularly known as the "stress tests," demonstrated that many institutions' information systems could not provide timely, accurate information about bank exposures to counterparties nor complete information about the risks posed by different positions and portfolios, regulators did not press firms vigorously enough to fix them (Bernanke, 2010). The Fed, in charge of regulating financial holding companies and state banks, did not identify and address abuses in subprime lending either.

None of this happened by chance. As Alan Greenspan, former Chairman of the Fed, recognized, "those of us who have looked to the self-interest of lending institutions to protect shareholders' equity, myself included, are in a state of shocked disbelief" (*New York Times*, 10/23/2008). The blind confidence in self-regulation through market forces was the belief behind the huge deregulating process that took place in the 1980s and 1990s in the US. The Gramm-Leach-Bliley Act removed barriers in the market among banking companies, securities companies and insurance companies, expressly recognized national banks' authority to engage directly in asset-backed securitization activities and repealed key provisions of the Glass-Steagall Act in order to allow banks to affiliate with full service investment banks that engage extensively in, among other securities activities, asset securitizations. This allowed national banks and companies affiliated with such banks to be fully involved in securitization activities.

However, not only the regulatory framework was weakened; the regulators' power was too. The tenor of the times was to keep regulation as low as possible. The FCIC quotes Richard Spillenkothen, the Fed's director of Banking Supervision and Regulation from 1991 to 2006, who discussed banking supervision in a memorandum submitted to the FCIC: "Supervisors understood that forceful and proactive supervision, especially early intervention before management weaknesses were reflected in poor financial performance, might be viewed as i) overly-intrusive, burdensome, and heavy-handed, ii) an undesirable constraint on credit availability, or iii) inconsistent with the Fed's public posture" (FCIC, 2011, 54).

The main concern was to create checks and balances and keep any agency from becoming arbitrary or inflexible. Hence the opposition to any initiative to consolidate bank regulation. The FCIC report quotes Alan Greenspan's 1994 testimony on this matter: "The current structure provides banks with a method … of shifting their regulator, an effective test that provides a limit on the arbitrary position or excessively rigid posture of any one regulator. The pressure of a potential loss of institutions has inhibited excessive regulation and acted as a countervailing force to the bias of a regulatory agency to overregulate" (Ibid.).

Under the Gramm-Leach-Bliley Act, the Fed supervised financial holding companies as a whole, looking only for risks that cut across the various subsidiaries owned by the holding

company. To avoid duplicating other regulators' work, the Fed was required to rely "to the fullest extent possible" on the examinations and reports of those agencies regarding subsidiaries of the holding company. According to the Fed's Chairman Ben Bernanke, this "made it difficult for any single regulator to reliably see the whole picture of activities and risks of large, complex banking institutions" (Ibid., 55).

Therefore, the financial regulatory system was deeply fragmented and weakened to avoid interference with the market wise behavior. However, some people and institutions warned about the risks at stake. For example, in 2002 the state of Georgia passed a law by which investment banks that created mortgage-backed securities would be liable for financial damage if mortgages turned out to be fraudulent. However, the OCC ruled that the Georgia law did not apply to national banks or their subsidiaries. Finally, the law was amended in 2003: the liability provision was curtailed and other elements of the law were eliminated (*Newsweek*, October 20, 2008).

When in 2004 the state of Michigan tried to examine the books of the mortgage unit of Wachovia Bank that operated in that state, the OCC denied authority to the states to intervene in the operations of national banks. Michigan claimed that the Constitution preserved the right of the states to protect their residents, but the Supreme Court ruled in April 2007 establishing that the OCC had exclusive powers over the bank. A year later, the Wachovia Bank had to be saved from bankruptcy through its acquisition by Wells Fargo.

Rajan's (2005) prescient analysis of how the developments observed in financial markets could degenerate into a crisis was not much listened to. No economic journal published his paper and on the SSRN site only collected 93 downloads, which made it rank **96,914[th]** on the SSRN download ranking.

Nouriel Roubini, Professor at New York University, in a presentation at the IMF in September 2006, predicted the outbreak of a crisis from a massive default in mortgages and the securities backed by them.[13] Nobody paid too much attention to his words, especially when they included the prognosis of massive bankruptcies of hedge funds, investment banks and other financial institutions such as Fannie Mae and Freddie Mac. A year and a half later, his predictions were fulfilled and Professor Roubini now travels the world giving talks explaining what happened and what is expected to happen.

However, these were isolated voices. A very typical argument in those Great Moderation days was the one reflected in the following quotation: "The passage of the Glass-Steagall Act was prompted by concerns about various kinds of abuses by commercial banks' investment banking affiliates, including overstating the quality of the underwritten securities issued by the commercial banks' clients, packaging bad commercial loans into securities, and misusing responsibility for trust accounts. Recent research, however, suggests that those concerns were invalid" (Kwan and Laderman, 1999, 18). Unfortunately, the 2007–08 events showed that the concerns that had prompted the 1933 Act were very well founded.

Even at the beginning of 2007, Wharton real estate professor Todd Sinai argued that three things had to happen for the subprime market to tank: borrowers' incomes had to drop,

---

[13] See http://www.economonitor.com/nouriel/2010/09/02/economonitor-flashback-roubinis-imf-speech-september-7-2006/.

interest rates had to rise and housing prices had to fall. "It is extremely rare that all three things happen," he said.[14]

The conclusion is that after the deregulation movement that took place during the 1980s and 1990s, the US financial regulatory system was unable to foresee, let alone prevent, the financial crisis.

2.6   Summary

Credit rating agencies played a decisive role in the development of the subprime mortgage meltdown. They were an essential input into the process of manufacturing vast quantities of triple-rated securities with attractive yields.

Banks pursued an aggressive lending policy in order to get rid of the excess money in their vaults. The rapid increase in market share by unregulated brokers and originators put pressure on regulated banks to lower their underwriting standards. Securities backed by subprime mortgages lent to borrowers whose abilities to repay were doubtful became prominent in banking businesses. Safe institutions that desired to be more careful and scrutinize more deeply the repayment capacities of their potential customers were afraid of losing market share and making fewer loans. Just as bad money drives out good, bad financial institutions could drive out good ones.

On top of this, a gigantic interlinked structure of securities was created with the help of the mostly unregulated shadow banking system. This structure was ready to amplify and spread to the whole financial system following the failure of any mortgage loan. The conditions were right for a perfect storm. When house prices finally stopped rising, borrowers could not refinance their way out of financial difficulty, mortgage defaults soared and the whole securitization building collapsed.

3   The European debt crisis

3.1   Introduction

In late 2009, the then recently appointed Greek Prime Minister George Papandreou announced that previous governments had failed to reveal the true size of the nation's deficits. Greece's debts were larger than had been reported.[15] After that, the Portuguese, Spanish and Italian public debts also became a matter of concern because their government debt/GDP ratios were near to the Greek one. The European sovereign debt crisis had started. This chapter is organized as follows. Section 3.2 analyzes the origin of the crisis in these European countries. In Section 3.3, the specificities of euro debt are discussed. Section 3.4 analyzes the case of Ireland whose debt crisis preceded the Greek one. Section 3.5 is devoted to the latter. The role of a single currency on regional imbalances is underlined in Section 3.6. The case of Spain is analyzed in Sections 3.7 and 3.8. Section 3.9 is devoted to the analysis of the Italian case. Section 3.10 summarizes the findings of the chapter.

---

[14] See http://knowledge.wharton.upenn.edu/article.cfm?articleid=1691.
[15] In fact, in 2004, Eurostat had already revealed that the statistics for the budget deficit had been under-reported at the time Greece was accepted into the European Monetary Union in 2000. According to Eurostat, the 1999 deficit was 3.4% of GDP instead of the originally reported 1.8%.

3.2 Evolution of countries' indebtedness

A first question has to do with the origin of the European debt crisis. Some people have pointed their fingers at the American financial crisis. "This crisis was not originated in Europe," claimed the EU Commission President Jose M. Barroso, who added: "This crisis originated in North America and much of our financial sector was contaminated by… unorthodox practices from some sectors of the financial market."[16]

However, as we shall see, Greece and Italy were already heavily indebted as early as 1996, long before the US financial crisis blew up. However, this does not exclude the possibility of some connection between both crises, which is explored below by comparing the debt situation before and after 2007.

A second question is how the debtor country governments as the Greek one became so highly indebted. A common explanation for this has been the following.[17]

Banks in Germany, France and elsewhere had bought and exposed themselves massively to Greek debt because they assumed that Greek debt, like other euro-area public debt, was essentially risk-free.

Because the monetary union made the commitment to low inflation more credible, the introduction of the euro in 2001 caused interest rates to fall in those countries where expectations of high inflation previously kept interest rates high.

Bond buyers assumed that a bond issued by any government in the European Monetary Union was equally safe. As a result, the interest rates on Greek and Italian government bonds were not significantly different from the interest rate on German government bonds. Governments responded to these low interest rates by increasing their borrowing.

However, the data do not endorse the former explanation. Table 3.1 shows the general government debt/GDP ratio in 2010 for those countries whose public debt ratio exceeded the average for the 27 EU countries as a whole. France and Germany are among the more than average indebted countries, which shows that high indebtedness is not solely a southern country phenomenon.

---

[16] *The Week.* June 20, 2012. http://theweek.com/article/index/229570/did-the-us-cause-the-european-debt-crisis.
[17] See, for example, Feldstein (2012).

Table 3.1 General government gross debt (percentage of GDP) - 2010

Country	2010
EU (27 countries)	80.1
Greece	144.9
Italy	118.4
Belgium	96.2
Portugal	93.3
Iceland	92.9
Ireland	92.5
Germany	83.2
France	82.3
Hungary	81.3

Source: Eurostat

Table 3.2 shows the evolution of government debt between 1996 and 2010 for a selected group of countries. First, it can be noted that some of the now highly indebted countries did not exceed the Maastricht limit of 60% of GDP until as recently as 2007.

Table 3.2 Evolution of general government gross debt (percentage of GDP) **-** 1996**/**2010

Country	2007	2008	2009	2010	2010/07
EU (27 countries)	59.00	62.5	74.7	80.1	35.76
Ireland	24.8	44.2	65.2	92.5	272.98
Iceland	28.5	70.3	87.9	92.9	225.96
Romania	12.8	13.4	23.6	31.0	142.19
UK	44.4	54.8	69.6	79.9	79.95
Spain	36.2	40.1	53.8	61.0	68.51
Portugal	68.3	71.6	83.0	93.3	36.60
Greece	107.4	113.0	129.3	144.9	34.92
Hungary	67.0	72.9	79.7	81.3	21.34
Italy	103.1	105.8	115.5	118.4	14.84
Belgium	84.1	89.3	95.9	96.2	14.39

Source: Eurostat

Second, the public debt to GDP ratios of Greece, Ireland, Belgium, Spain and Italy were almost the same in 2007 as they were in 2001 (in some cases, they were even lower). This contradicts the idea that it was the introduction of the euro and the consequent fall in interest rates that stimulated governments to substantially increase their borrowing.

On the other hand, Greece, Italy, Portugal, Belgium and Hungary had already exceeded the 60% Maastricht limit in 2007,[18] when the American subprime crisis started. However, they shared the slowest increasing government debt/GDP ratios between 2007 and 2010. Even more, by 1996 – before the introduction of the euro– Italy, Greece and Belgium were already highly indebted countries.

---

[18] As Hungary is not a member of the euro-zone, the Maastricht criteria was not mandatory for it.

Therefore, we can distinguish a first group of countries whose debt problems have roots before 2007 and did not worsen significantly after that year: Greece, Italy, Portugal, Belgium and Hungary. Moreover, by 2001 Greece's public debt/GDP ratio was already 103.7 compared with 108.2 for Italy and 106.5 for Belgium. This last country is a special case because it is the only one in the group that reduced its debt between 2001 and 2007.

A second group is formed by those "new" highly indebted countries: Ireland and Iceland. They showed the highest rates of increase in their public debt to GDP ratios between 2007 and 2010 and their 2010 ratios were above the average for the EU. Romania also had a fast growing ratio but the level of public debt attained in 2010 as a percentage of GDP was still far below the average for the EU.

The United Kingdom comes immediately below these countries with a debt to GDP ratio practically equivalent to the EU average. Finally, we have Spain, whose government debt to GDP ratio was in 2010 only a bit above the Maastricht limit and had increased at a lower rate than the UK's ratio between 2007 and 2010. However, while the UK's debt was considered to be safe, Spain's debt was no better rated than those of Portugal or Italy.

Thus, there are different cases to consider rather than a single story for European countries' indebtedness process. The idea that we may have a unique explanation for the debt crisis is also presented in Perez-Caldentey and Vernengo (2012, 3), who argue that "the crisis in Europe is the result of an imbalance between core and noncore countries that is inherent in the euro economic model." They also maintain that it was the euro, and its effects on external competitiveness, that triggered mounting disequilibria and debt accumulation in noncore countries or peripheries. As we will see, this argument seems to be valid to a certain extent just in the cases of Greece and Portugal, but not for the rest of the countries involved in the crisis where other factors seem to have played a major role.

In what follows, we concentrate our analysis on the five euro-area countries in the eye of the debt crisis storm with a casual reference to the case of Iceland.[19]

3.3   Specificities of the euro-area public debt

A first peculiarity of the euro-area public debt is that, strictly speaking, it is neither purely domestic nor purely external. Most of the public debt issued by euro-area countries is denominated in euro and is mostly held by euro-area residents. Yet, it is different from the domestic debt of countries owning their own currencies because more of it is held outside the issuing country and because the issuing country does not have full control over the currency in which the debt is denominated. Therefore, debt in the euro-area can be considered to be both 'foreign' and 'domestic' (Gianviti et al., 2010, 18).

This means that euro-area public debt is not subject to the currency mismatch associated with external debt: governments have to pay their debts in the same currency they collect their revenues. However, it also means that a national government cannot revert to high inflation to rid itself of an excessive debt burden, as might be the case if the debt were strictly domestic. The European Monetary Union seems to assume that sovereign debt crises cannot happen. At least, it has no provision for them. Moreover, the common reading of Article 125 of the Lisbon Treaty has been that it rules out the possibility of a bailout of an EU member state by

---

[19] The Cyprus banking crisis is an especial case, mainly the result of the Greek sovereign debt haircut, although it has something in common with Iceland´s case.

other member states or by the EU. Therefore, without these inflation and bailout channels, a country with a situation of excessive debt has only two ways out of it: severe and harmful fiscal retrenchment or default.

3.4   The new highly indebted countries: the case of Ireland

Ireland's economy had by 2007 already become dangerously dependent on construction and housing as a source of economic growth and tax revenue. A lightly regulated financial system fed on this process. In fact, the growing construction boom was fuelled by the increasing reliance of Irish banks on wholesale external borrowing at a time when international financial markets were awash with cheap investable funds. The fact that Ireland was a founder member of the euro-zone brought a dramatic and sustained fall in nominal and real interest rates that stimulated the protracted building boom. Specific tax incentives boosted the overheated construction sector. From late 2003 onwards, banks stimulated demand with financial innovations such as 100% loan-to-value mortgages.

When the global economic environment changed at the beginning of 2007, Irish residential property prices started falling and kept falling during the rest of 2007 and 2008. Heavy loan losses on the development property portfolios acquired at the peak of the market became inevitable. The decline in property prices and the collapse in construction activity resulted in severe losses in the Irish banking system. The story is not very different from the one that led to the US subprime crisis. "In their anxiety to protect market share against the competitive inroads of Anglo Irish Bank and UK-based retail lenders, their (Irish) banks' management tolerated a gradual lowering of lending standards, including decisions to authorize numerous exceptions to stated policies." (Governor of the Central Bank of Ireland, 2010, 8). This was tolerated by an unduly deferential approach to the banking industry by regulators. Outside bodies such as the IMF and OECD never drew attention to the threats that lay ahead.

Although banks carried out a quantification of risks in the context of the stress test exercises reported annually to the regulatory authority, "the capacity of the banks to undertake the exercise differed greatly; indeed none of them had reliable models, tested and calibrated on Irish data, which could credibly predict loan losses under varying scenarios" (Ibid., 11).

While at the end of 2003, the net indebtedness of Irish banks to the rest of the world was just 10% of GDP, by early 2008 borrowing, mainly for property, had jumped to over 60% of GDP. By early 2008, Irish banks found it more difficult to maintain funding in the international wholesale markets and, at the same time, there was a more rapid pull back by domestic investors from the property market.

Two weeks after Lehman Brothers announced it would file for Chapter 11 bankruptcy protection, the provision of a blanket system-wide state guarantee for Irish banks was announced. This measure was taken because of the drain of liquidity that had been affecting all Irish banks and that had brought one important bank to the point of failure.

Government spending doubled in real terms between 1995 and 2007, rising at an annual average rate of 6%. With the economy growing at an even faster rate, this implied a generally falling or stable expenditure ratio of expenditure to GDP until 2003. However, thereafter the ratio rose, especially after output growth began to slow in 2007 and the collapse in tax revenues in 2008–09. Much of the reason for the revenue collapse lies in the systematic shift over the previous two decades away from stable and reliable sources such as personal

income tax, VAT and excises towards cyclically sensitive taxes as corporation tax, stamp duties and capital gains tax.

In April 2009, the Irish government established the National Asset Management Agency (NAMA), with the mandate to purchase the universe of development-related loans (above a certain value) from banks. This category of loans was the main source of uncertainty concerning total loan losses. During 2009–10, NAMA purchased most of these loans at a steep average discount, but this meant that banks required substantial upfront recapitalization programs, which could only be provided by the state. These higher capitalization costs led to a sharp increase in gross government debt. Extra capital requirements by the banking system in 2009 and 2010 contributed to increased market concerns about the sustainability of the fiscal position. In fact, the deficit, as measured by the general government balance, widened from balance in 2007 to 7.3% of GDP in 2008 and to 14.1% in 2009, before it increased to 31.2% of GDP in 2010 due to the substantial government support to Irish banks. Excluding support to the banking system, the deficit was 11.5% of GDP in 2009 and 10.9% of GDP in 2010. The public funds aimed at rescuing the Irish banking sector represented 12.5% of Ireland's GDP. As shown in Table 3.2, Irish public debt soared from 24.8% of GDP in 2007 to 92.5% in 2010. Finally, the Irish government had to request assistance from the EU and IMF in November 2010 to avoid default on its public debt.

The case of Iceland

Although it has many features in common with the Irish one, Iceland's case has some particularities. The first one is that Iceland does not belong to the euro-zone. Property lending was neither as central to the Icelandic case. Access to international financial markets was, for banks, the principal premise for their large growth. Because of their – at that time – good credit rating, they had access to European markets; when funding in European debt securities markets became more difficult, the debt securities market in the US opened up. That opening was largely due to CDOs. Icelandic bank securities were packaged into these CDOs because of the high credit rating of the Icelandic financial undertakings, according to rating agencies. Further, Icelandic banks paid high interest rates considering that credit rating.

Thanks to the injection of foreign funds, the Icelandic financial system became far too large relative to the size of the Icelandic economy. On the other hand, the largest owners of all the large banks had abnormally easy access to credit at the banks they owned. The examination conducted by the Icelandic Special Investigation Commission showed that in the three largest banks, their principal owners were among the largest borrowers. The money market funds under the aegis of the management companies of these banks invested a great deal in securities connected to the owners of the banks.

Bank risk was highly concentrated. This applied both to lending to certain groups within each bank as well as to how the same groups also constituted high-risk exposures in more than one bank. Moreover, the banks had invested funds equivalent to more than 25% of their capital bases in their own shares. In addition, each of them invested in other banks' shares. It seems that the financing of owners' equity in the Icelandic banking system had been based, to such a great extent, on borrowing from the system itself. The shares owned by the largest shareholders of the banks were especially leveraged.

The onset of the international financial crisis in 2007 found Icelandic banks increasingly dependent on funding through international financial markets. Total deposits in the banks kept

shrinking from the autumn of 2007 until their collapse. Collateralized loans, mostly from the Central Bank of Ireland and the European Central Bank (ECB), increased substantially in all three banks as the liquidity crisis became more widespread.

When the prices of shares started dropping, all banks purchased their own shares on a large scale. As stated before, the banks held a lot of their own shares as collateral for their lending. With share prices declining, the quality of their loan portfolios would decline. Finally, the Financial Supervisory Authority of Iceland took over the domestic operations of the three largest banks in October 2008.

Outside Iceland, more than half a million depositors (far more than the entire population of Iceland) found their bank accounts frozen when the banks finally collapsed. In August 2009, a bill was passed to pay the United Kingdom and the Netherlands more than $5 billion lost in Icelandic deposit accounts. The Icelandic government debt increased from 28.5% of GDP in 2007 to 70.3% in 2008 after the takeover of the three largest Icelandic banks.

3.5   The "old" indebted countries: the case of Greece

As stated before, Greece did not comply with the Maastricht criterion with respect to the budget deficit at the time it joined the euro-zone in 2001. "Creative" statistics allowed it to be admitted into what has been conceived as a very exclusive club. Its debt/GDP ratio was already 103.7 in 2001, far above the 60% Maastricht criterion.[20] However, it declined to 97.4 in 2003. From then on, it kept increasing until reaching 144.9 in 2010. This reflected the increasing budget deficit Greece's public accounts had shown since 2000 (Table III.3).

Table 3.3 General government expenditure, revenue and deficit 2000/11 (percentage of GDP)

Year	Expenditure	Revenue	Balance
2000	46.7	43.0	-3.7
2001	45.3	40.9	-4.4
2002	45.1	40.3	-4.8
2003	44.7	39.0	-5.7
2004	45.5	38.1	-7.4
2005	44.6	39.0	-5.6
2006	45.2	39.2	-6.0
2007	47.6	40.8	-6.8
2008	50.6	40.7	-9.9
2009	53.8	38.2	-15.6
2010	50.2	39.7	-10.5
2011	50.1	40.9	-9.2

Source: Eurostat

---

[20] Notwithstanding its noncompliance with the Maastricht debt standard, Greece was admitted with the argument that it was expected to be making progress over time towards that goal.

Entrance into the euro-zone meant that Greece –as the other members of the euro-zone- gave up one of the tools a country has to reduce its budget deficit: devaluation. In fact, in equilibrium:

$$(I_d - S) + (G - T) = M - X$$

where I_d is domestic investment, S is national saving, G is government expenditure, T is government revenue and (M – X) stands for current account balance. A devaluation will reduce the value of (M – X); if the domestic private balance does not change, the government balance will be reduced.[21] The most direct way to do this is by taxing exports, as Argentina did in 2002, where export taxes absorbed a good part of the devaluation effect on exportable domestic prices.

As a matter of fact, Georgantopoulos and Tsamis (2011, 161) find for Greece, during the period 1980–2009, a significant unidirectional causal relationship between exchange rates and budget deficit running from the nominal effective exchange rate to the budget deficit. Moreover, they concluded that "a significant part of budget deficits' variance is caused by exchange rates since with a seven period lag 61.89% of [the budget deficit] is explained by [the nominal effective exchange rate] and by the end of the ten-year lag 83.97% of budget deficits' variance is caused by nominal effective exchange rates."

The continuous revaluation of the euro worsened Greece's budget imbalance after 2000. Figure 3.1 illustrates the relationship between the euro/dollar rate of exchange and the one-year lagged budget deficit/GDP ratio between 2000 and 2011. This runs in the same direction as the relationship found by Georgantopoulos and Tsamis.

Figure 3.1 Budget deficit and euro rate of exchange 2000–2011

---
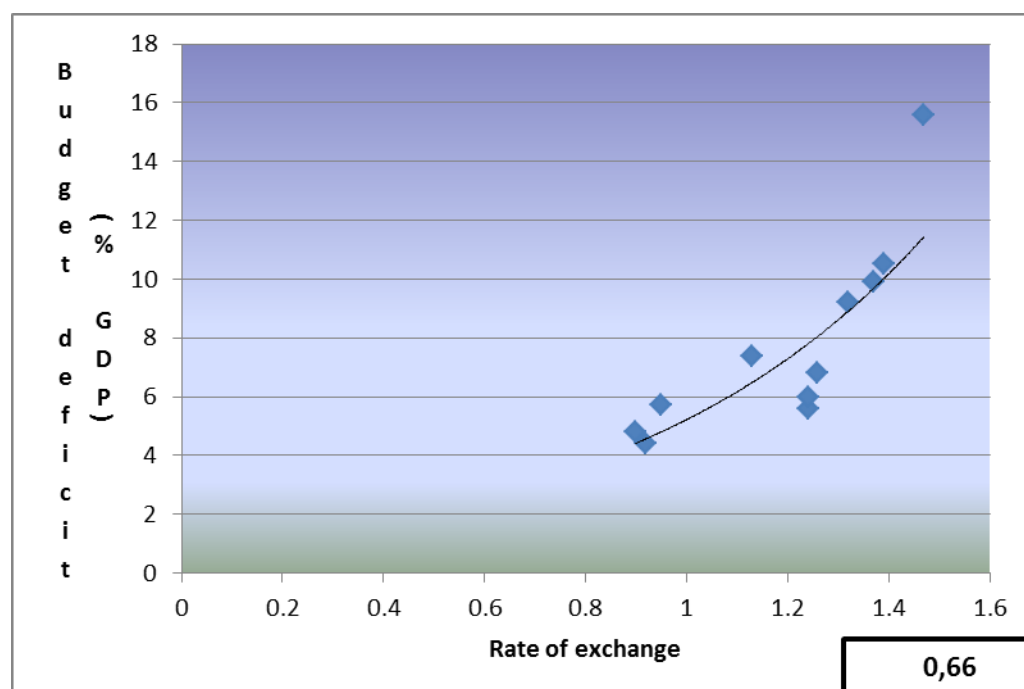
[21] The opposite happens, of course, in the case of a revaluation of the local currency.

What is the explanation for this positive association between the rate of exchange and budget imbalance? The appreciation of the euro[22] resulted in a loss of external competitiveness in the Greek economy, which led to a persistent deficit in the current account (Figure 3.2). An appreciation of the real exchange rate increases the purchasing power of domestic incomes in terms of imported goods. More imports and fewer exports result in a slowdown in economic activity. Tax revenues decline, while the government feels compelled to keep or increase public expenditure to make up for the decline in private demand. The budget deficit increases and so does public debt. Increasing demand for funds by the public sector leads to an increase in interest rates, which depresses again economic activity. According to the figures in Table III.3, public revenues have declined since Greece joined the euro-zone; since 2007, public expenditure increased, accelerating the rise in the budget deficit.

Figure 3.2 Current account deficit and the euro rate of exchange 2001/11
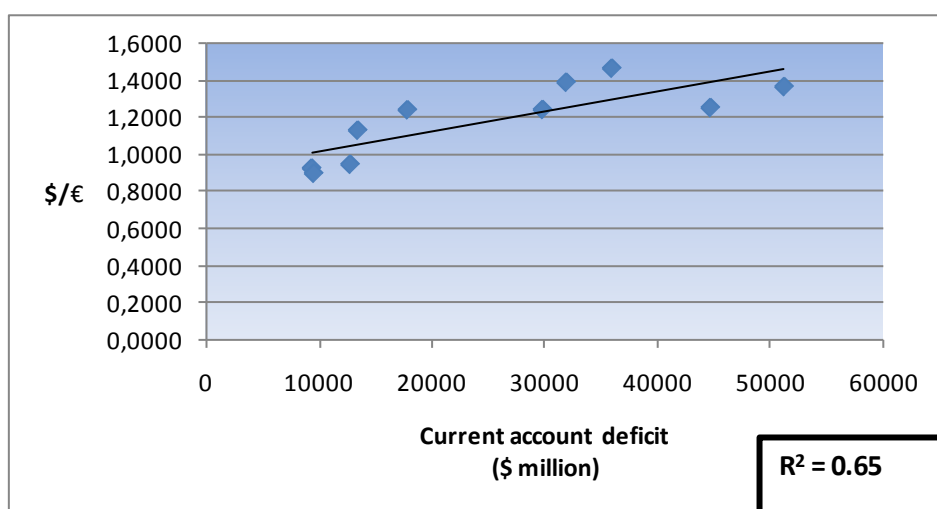
However, in the literature related to the "twin deficits hypothesis," it has usually being argued that causality runs from the government budget deficit to the current account, not the other way around. However, empirical studies are far from conclusive: in some cases, they support the conventional hypothesis;[23] others support the reverse causality running from the current account deficit to the fiscal deficit;[24] some support the Ricardian equivalence that budget and trade deficits are not correlated.[25] And, finally, some find both types of evidence or a bilateral relationship.[26]

In the case of Greece, it is clear that, since the introduction of the euro, causality cannot run from the budget deficit to the nominal rate of exchange. Moreover, when the budget deficit variable is introduced with a one-year lag.

---

[22] The exchange rate between dollar and euro was, in October 2000, 0.85 $/€ and reached in April 2008, 1.60 $/€, an appreciation of 88%.
[23] Abell (1990), Bachman (1992), Piersanti (2000), Leachman and Francis (2002), Cavallo (2005) and Erceg et al. (2005).
[24] Anoruo and Ramchander (1998), Khalid and Teo (1999) and Alkswani (2000).
[25] Miller and Russek (1989), Dewald and Ulan (1990), Enders and Lee (1990) and Kim (1995).
[26] Mukhtar et al. (2007) and Islam (1998).

The increasing Greek debt was primarily the result of growing budget deficits triggered by the appreciation of the euro and the consequent loss of competitiveness experienced by the Greek economy. This brings us to the issue of regional imbalances raised by Perez-Caldentey and Vernengo (2012).

3.6 The exchange rate and regional imbalances

The euro-area aggregate trade and current account position have always been close to balance but this only means that the euro rate of exchange is in line with the competitiveness of the core countries of the euro-zone. Many industries in Greece and other peripheral countries are not competitive at that rate of exchange; that is why these countries run increasing current account deficits (see Table III.4). In fact, external imbalances diverge sharply in the euro-area: while Germany, the Netherlands and Finland run significant surpluses, countries in southern Europe run huge deficits. By the way, it is worthwhile noting that Germany had run persistent current account deficits during the nineties which turned into surpluses only after 2000.

Table 3.4 Current account balance in selected EMU countries- 2001/10 (Percentage of GDP)

	2001	2002	2003	2004	2005
France	1.8	1.2	0.7	0.5	-0.5
Germany	0.0	2.0	1.9	4.7	5.1
Netherlands	2.6	2.6	5.5	7.6	7.4
Finland	8.4	8.5	4.8	6.2	3.4
Greece	-7.2	-6.5	-6.5	-5.8	-7.6
Italy	0.3	-0.4	-0.8	-0.3	-0.9
Portugal	-10.3	-8.2	-6.4	-8.3	-10.3
Spain	-3.9	-3.3	-3.5	-5.2	-7.4

	2006	2007	2008	2009	2010
France	-0.6	-1.0	-1.7	-1.5	-1.7
Germany	6.3	7.5	6.3	5.6	5.7
Netherlands	9.3	6.7	4.3	4.2	6.6
Finland	4.2	4.3	2.6	1.8	1.4
Greece	-11.4	-14.6	-14.9	-11.1	-10.1
Italy	-1.5	-1.3	-2.9	-2.0	-3.5
Portugal	-10.7	-10.1	-12.6	-10.9	-10.0
Spain	-9.0	-10.0	-9.6	-5.2	-4.6

Source: Eurostat

The euro-zone reproduces the sort of regional problems that exist within many countries. There is a highly competitive core and a relatively backward periphery. Therefore, a long-run strategy for regional convergence is needed and, at the same time, a short-run one to smooth

the transition process. Although EU regional policy aims at promoting the "harmonious, balanced and sustainable development of the European Union," it has proven up to now to be insufficient to face the specific consequences of the monetary union. Therefore, the Greek government had to face the outcome of joining the euro-zone and had to take decisions that resulted in a worsening of the heavy indebtedness pre-existing at the time of joining the euro-zone.

Katsimi and Moutos (2010) emphasise the role of current of account imbalances due to the loss in Greek international competitiveness. However, productivity gaps and external deficits exist within each country. All American states have the same productivity? What about East and West Germany? Who cares what their external balances are? A region within a country can run a current account deficit indefinitely as long as there is a transfer of resources from the richer to the poorer regions. Therefore, this should not be a problem for the eurozone provided those who, thanks to the euro-zone, benefit of external surpluses are ready to transfer resources to the backward periphery. This is the real issue at stake as far as the productivity gap is concerned.

Germany's unification process could have been an interesting antecedent to take into consideration. The major economic implication of German economic and monetary union was precisely that East Germany would run a current account deficit with the rest of the country that was financed by transfers from the West. In the case of Germany, the New Länder began with an enormous competitive disadvantage and West Germans were supposed to transfer between 3% and 4% of GDP per annum to the East (Carlin, 1998, 16). However, no provision was taken in the euro-zone to make up for the short-run negative consequences that peripheral economies could suffer from joining the euro.[27]

In fact, when the monetary union was implemented in 1999, the functioning of the single currency was seen as a sort of panacea, making additional policy targeting seem superfluous. However, the result has been an increasing current account deficit for Greece and other peripheral countries. What has not been done before in the form of resource transfers from the richer to the poorer countries of the euro-zone has to be done in the way of helping these countries restructure their debts.

Somebody may argue that internal devaluation is the way through which Greek could become competitive. Downwards price and wage inflexibility makes this a very painful and unbearably long process. It did not work in Argentina, which, after three years of an ever-deepening recession/depression, had no alternative but to default and devalue its currency. It does not seem to be a valid alternative for Greece either.

The success of the 2012 Greek restructuring makes it more likely that debt restructuring will be seriously considered as a policy option if additional European countries lose market access, as Zettelmeyer et al. (2012) point out.

---

[27] I refer here to the specific consequences of joining the euro, which are independent of those following the EU integration to make up for which there were significant resource transfers, particularly through structural funds.

The case of Portugal

In the second half of the 1990s, Portugal showed impressive economic results. Its GDP per capita grew faster than the EU average and Portugal fulfilled the Maastricht criteria for the monetary union. However, by 2000 Portugal had already become the first country to be subjected to the EU's Excessive Deficit Procedure specified in the Stability and Growth Pact legislation, and again in 2005 when its deficit reached more than 6%.[28]

As in the case of Greece, the continuous revaluation of the euro worsened Portugal's budget imbalance after 2000. Figure 3.3 illustrates the positive relationship between the euro/dollar rate of exchange and the one-year lagged budget deficit /GDP ratio between 2001 and 2011.

Figure 3.3 Portugal's budget deficit and the euro/dollar rate of exchange 2001/11
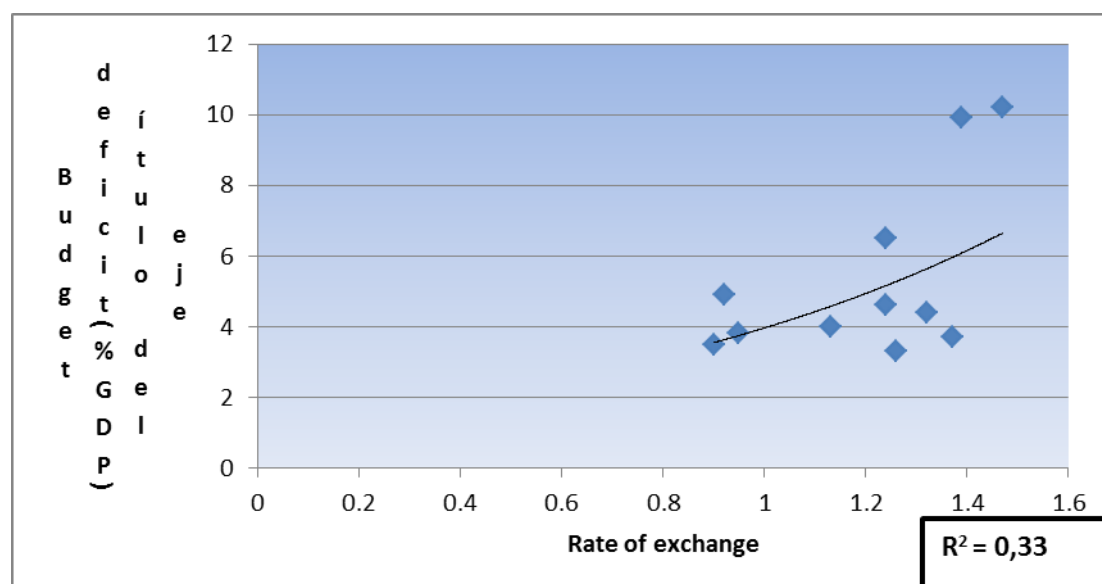
However, the financial crisis worsened Portugal's economic situation. Its impact was first felt in Portugal at the beginning of 2008, with a severe credit squeeze, a reduction in banks' abilities to access capital markets and the collapse of two banks: BPN, which was nationalized in November 2008, and BPP, which was intervened in by the state and finally went bankrupt in 2010. The Portuguese government reacted by implementing an "Initiative to Strengthen Financial Stability," which focused on improving the information and transparency obligations of financial institutions, increasing deposit guarantees, granting state guarantees to banks and strengthening their financial soundness. These measures – particularly the nationalization of BPN and the intervention in BPP – implied an increase in public deficit and public debt. The international financial crisis, shrinking exports, declining investment (including in construction) and dampened consumer spending all contributed to the contraction of Portugal's economy.

---

[28] Indeed, throughout the entire democratic period following the 1974 revolution, Portugal never had a surplus in the state budget.

Table 3.5 Annual rates of growth 2000–11

Year	GDP rate of growth
2000	3.916
2001	1.975
2002	0.764
2003	-0.911
2004	1.560
2005	0.775
2006	1.448
2007	2.365
2008	-0.009
2009	-2.908
2010	1.401
2011	-1.669

Source: Eurostat

Portugal's already low rate of growth became negative in 2008 and 2009. The first reaction to the crisis was to stimulate demand. This increase in public expenditure on top of the measures taken to preserve the Portuguese financial system meant that the public deficit soared to 10.2% in 2009 and Portugal's public debt to GDP ratio jumped from 68.3% in 2007 to 93.3% in 2010. However, public accounts improved in 2011 after a series of tax hikes and salary cuts for public servants took place.

Table 3.6 General government expenditure, revenue and balance 2000/11 (Percentage of GDP)

Year	Expenditure	Revenue	Balance
2000	41,6	38,3	-3,3
2001	43,2	38,3	-4,9
2002	43,1	39,6	-3,5
2003	44,7	40,9	-3,8
2004	45,4	41,4	-4,0
2005	46,6	40,1	-6,5
2006	45,2	40,6	-4,6
2007	44,4	41,1	-3,3
2008	44,8	41,1	-3,7
2009	49,8	39,6	-10,2
2010	51,3	41,4	-9,9
2011	49,4	45,0	-4,4

Source: Eurostat

These measures allowed Portugal, in the first half of 2011, to receive a €78 billion IMF/EU bailout package in a bid to stabilize its public finances, as Greece and Ireland had done before. In 2012, the Portuguese government used €3 billion from the bailout package to rescue Portugal's largest listed bank by assets, Millennium BCP. By the end of 2012, Portugal had regained access to financial markets when the state managed to renew one-third of the outstanding bonds at a reasonable yield level (5.12%). The bailout funding program was supposed to run until June 2014, but at the same time it requires Portugal to regain complete bond market access by September 2013. While the budget deficit for 2012 was forecasted to end at 5%, the country is expected to reduce the budget deficit to a level below 3% of GDP in 2014.

3.7   Spain: a special case

The weight of Spain's public debt as of 2011 was substantially lower than the weight of the debt of the United Kingdom and of Germany. Spain's government debt ratio was just 68.5 of GDP against 85.7 in the UK and 81.2 in Germany, not to mention 165.3 in Greece and 120.1 in Italy. Why was, then, Spain involved in the European financial crisis? There is just one single reason: because it evoked the Irish case. In 2007, the public debt to GDP ratio in Ireland was only 24.8. However, it soared to 65.2 in 2009.

As in Ireland, construction had been a fast growing industry in Spain. It expanded at a rate of 5% per year between 1996 and 2007. Between 1998 and 2007, the number of housing units grew 30% (Arellano and Bentolila, 2009, 28). House prices increased dramatically and people expected the process to go on without an end. Real house prices – house prices adjusted for the change in the consumer price index – increased by 127% between 1996 and 2007 (André, 2010, 9). Therefore, real estate became the preferred destination for savings. Tax benefits[29] stimulated even greater demand for real estate, biasing household investment to housing in place of other types of assets. This process was reinforced after 1999. After becoming a member of the euro-zone, Spain benefited – as in the case of Greece and other southern Europe countries – from a drastic reduction in interest rates. The flight of capital from the equity markets that occurred between 2000 and 2003 was primarily funneled to the real estate sector. Loans became available at lower interest rates. Therefore, businesses and individuals saw their borrowing capacities increase; this stimulated the demand for house building. Housing became a shelter for assets: real estate investments promised attractive capital gains. Houses were bought because prices were expected to rise and prices rose because there were more and more purchases increasingly financed by loans. The construction market flourished. Banks offered 40-year and, later, even 50-year mortgages. The construction sector increased its share of Spanish GDP from 6.9% in 1995 to a high of 10.8% in 2006. In 2007, construction accounted for 13.3% of total employment. However, that year, coinciding with the global economic crisis, the real estate bubble burst. When international liquidity – until then cheap and plentiful – started lacking, the Spanish real estate market entered a crisis. Prices started declining in 2008.

Regional loans and savings banks, the so-called "*cajas*," were very active in the real estate market. They owned 56% of the country's mortgages in 2009. They were the first victims when the market crashed that year: debtors fell into bankruptcy and bad loans dramatically increased. In March 2009, the Spanish government announced its first bailout of a *caja*. After that, more bank bailouts were announced by the Spanish government. While these

---

[29] Altogether, 15% of mortgage payments are deductible from personal income taxes in Spain.

government bailouts kept these banks from going bankrupt, investor confidence in the Spanish economy sunk even lower. Many real estate developers avoided bankruptcy only because banks kept permitting them to refinance their loans. In this way, loans were reported as performing. In May 2012, Bankia, a bank that resulted from the merger of several *cajas*, had to be bailed out by the government. At that time, it was the fourth bank by size in the Spanish ranking of banking institutions.

3.8   The evolution of public finance in Spain

Table 3.7 shows the evolution of general government expenditure, revenue and deficit between 2000 and 2011. It shows that Spain had a small deficit between 2000 and 2004, far below the ceiling of 3% of GDP that the European Stability and Growth Pact established for member states after the introduction of the euro on January 1, 1999. From 2005 to 2007, the increase in revenues allowed the government to run a surplus. The situation abruptly reversed in 2008 precipitated by a significant decrease in revenues, a decline that deepened in the following years, as a reflection of the international financial crisis.

Table 3.7 General government expenditure, revenue and balance 2000/11 (Percentage of GDP)

Year	Expenditure	Revenue	Balance
2000	39.2	38.2	-0.9
2001	38.7	38.1	-0.5
2002	38.9	38.7	-0.2
2003	38.4	38.0	-0.3
2004	38.9	38.8	-0.1
2005	38.4	39.7	1.3
2006	38.4	40.7	2.4
2007	39.2	41.1	1.9
2008	41.5	37.0	-4.5
2009	46.3	35.1	-11.2
2010	45.6	36.3	-9.3
2011	43.6	35.1	-8.5

Source: Eurostat

As can be seen in Table 3.8, the rate of growth plummeted in 2008 and became negative in 2009 and 2010. The contraction in international liquidity supply was followed by a restriction on credit and subsequently by a sharp decline in construction and employment. The increase in unemployment meant a rise in spending on unemployment and other social benefits. The bailout of several *cajas* was another source of increase in public expenditure. On the other hand, the decline in GDP was followed by a weakening of public revenues, especially those linked with the real estate sector.

Table 3.8 Annual rates of growth 2000/11

Year	GDP rate of growth
2000	5.00%
2001	3.60%
2002	2.70%
2003	3.10%
2004	3.30%
2005	3.60%
2006	4.00%
2007	3.60%
2008	0.90%
2009	-3.70%
2010	-0.10%
2011	0.70%

Source: INE

Therefore, the swift deterioration of Spain's public finance flashed warning lights on the capacity of its government to face the services of its increasing public debt, which had exceptionally short maturity structures. Spain was following Ireland's steps with a three-year delay.

3.9   Italy: a different "old" debtor

The Italian government was highly indebted long before the crisis outburst. In 2007, the general government debt to GDP ratio was already 103.1, second only to Greece, and well above the 60% Maastrict criterion. However, nobody worried at that time for the Italian public debt and the Italian government had no problem refinancing it. Between 2007 and 2010, it only increased 15%.

However, the American financial crisis deeply affected the Italian economy. The transmission mechanism was the contraction in the interbank loan market that was the immediate consequence of the crisis. Banks refused to lend money to each other because of a lack of liquidity and the uncertainty about the financial soundness of borrowers. Besides the contraction in liquidity, Italian banks were also affected by their close links with central and eastern European countries where they had built a network of branches and affiliated banks. There was a risk of the collapse or illiquidity of this part of the network. The government responded to the risk of banking crisis by guaranteeing bank deposits to a maximum of €103,000 in the event of a bankruptcy. This avoided a bank run on deposits. However, banks reacted to the liquidity crisis by reducing credit to clients and consumers and raising the amount of collateral required for new loans. These measures affected investment and consumption. Bugamelli et al. (2009, 11) estimate that in the period from January 2008 to June 2009 production fell by more than 35% in sectors such as electrical machinery,

metallurgy and cars. The GDP rate of growth became negative in 2008 and 2009 (Table III.9). Growth resumed in 2010, but was snuffed out in 2011.

Table 3.9 Annual rates of growth 2000/11

Year	%
2000	3.7
2001	1.9
2002	0.5
2003	0.0
2004	1.7
2005	0.9
2006	2.2
2007	1.7
2008	-1.2
2009	-5.5
2010	1.8
2011	0.4

Source: Eurostat

The reduction in economic activity cut the amount of tax collected and anti-cyclical policies increased public expenditure. As a result, there was a significant increase in the public deficit (see Table 3.10).

Table 3.10 General government balance 2000/11 (Percentage of GDP)

Year	%
2000	0.8
2001	-3.1
2002	-3.1
2003	-3.6
2004	-3.5
2005	-4.4
2006	-3.4
2007	-1.6
2008	-2.7
2009	-5.4
2010	-4.6
2011	-3.9

Source: Eurostat

After Berlusconi stepped down, the new Prime Minister Mario Monti launched a deep austerity plan including measures such as increasing the retirement age, raising property taxes, simplifying the operation of government agencies and going after tax evaders.

In contrast to most European countries, the banking system in Italy practically did not resort to any public help between 2008 and 2011. Italian banks mainly faced the crisis by raising funds in capital markets. Italy's banking system required very low support from the ECB (Table III.11). The results of the EU-wide stress test carried out by the European Banking Association in 2010 and 2011 show that the included Italian banks successfully passed the test. Moreover, the Italian banking system seems to have low exposure to government debt; it holds less than 10% of domestic public debt –against more than 40% in the case of Spanish banks – as well as low exposure to foreign sovereign risk, which represents only 23% of the total government debt Italian banks hold (see Bolton and Jeanne, 2011).

Table 3.11 Funds provided by the ECB to national banking systems as of December 2011 Percentage of GDP

Country	%
Ireland	87.79
Greece	61.46
Portugal	27.65
Netherlands	26.9
Spain	16.83
Italy	12.65
France	10.89
Belgium	9.54
Austria	4.5
Germany	2.16

Source: OECD

Therefore, in contrast to Spain, Italy's problem seems to be essentially located in its public debt, whose ratio to GDP, although high, is no worse than it was 20 years ago, when nobody worried about it. In fact, the country's debt first hit 120% of GDP in 1993, after the public deficit reached 9.5% of GDP in 1992.

After the exchange rate turmoil that hit the European monetary system in 1992, Italy devalued the lira. Italian trade performance improved as import growth slowed, while export growth remained relatively constant. Therefore, Italy went into the euro-zone with a large surplus on its trade accounts. The high levels of Italian public debt only became a problem when, in the context of the 2011/12 European economic climate, the private sector began to lose confidence in the ability of the Italian state to service its debt.

3.10   Summary

The European indebtedness process does not accept a unique explanation. Of course, it may be argued that the European as well as the American crises are just chapters in a *global* credit bubble (McKinsey Global Institute, 2011) or the consequences of a global money or

savings glut. However, this explains little except that Europeans and Americans have had access to cheap money during the past 10 years.

This paper shows that among the most indebted European countries there are at least two different groups. One made up of "old" debtors, whose debt to GDP ratios slightly grew between 2001 and 2007. This means that in these countries the debt problem antecedes the introduction of the euro. A second group of "new" debtors comprises those countries whose debt suddenly increased as a result of the 2007/08 financial crisis. These are the cases of Ireland and Iceland.

Spain is a special case whose debt to GDP ratio was substantially lower than the weight of the debt of the United Kingdom and Germany not to mention Greece or Italy. However, its public debt was severely punished by the market because of the doubts about its banking system's health, which raised suspicion that it might require governmental support, as in the cases of Ireland and Iceland.

Therefore, although it is true that the US financial crisis triggered the European debt crisis, it did it through different channels. In the cases of Ireland and Iceland, through a severe credit squeeze and a reduction in banks' abilities to access the capital markets. The drain of liquidity experienced by the banking system precipitated governmental intervention with the consequential jump in public debt. However, in the cases of Greece, Italy and Portugal, the American financial crisis mainly brought attention upon the fiscal situation of countries already heavily indebted, who could face growing difficulties to roll over their debts in an increasing climate of fear and distrust.

Far from helping to reverse their pre-existing fiscal imbalances, entrance into the euro-zone had aggravated them for Greece and Portugal. In fact, the continuous revaluation of the euro worsened their budget imbalances after 2000, increasing their public debt. A positive association between the rate of exchange and budget imbalance was found for both countries. After the debt crisis burst, both countries found themselves without access to capital markets and had to resort to IMF/EU bailout packages in an attempt to stabilize their public finances.

In 2007, Italy's general government debt to GDP ratio was 103.1, second only to Greece, and well above the 60% Maastricht criterion. However, nobody worried at that time for the Italian public debt and the Italian government had no problem in refinancing it. Moreover, it only increased 15% between 2007 and 2010. Therefore, the Italian debt crisis is a clear example of the change in humor in financial markets after the American financial crisis.

The announcement by the President of the ECB, in mid-2012, that the ECB would become the euro-zone's lender of last resort by starting to purchase the sovereign bonds of the area's stricken economies calmed the waters, allowing European authorities to buy time to figure out how they could get the area out of the debt crisis.

As Lane (2012, 60) points out, a country with a high level of sovereign debt is vulnerable to increases in the interest rate. "This risk can give rise to self-fulfilling speculative attacks: an increase in perceptions of default risk induces investors to demand higher yields, which in turn makes default more likely." The opposite happens if default risk is perceived to be low. So, we are in the presence of a multiple equilibria problem. The announcement by the ECB acted as a signal to push the system to the "good" equilibrium.

On top of this, a new European Stability Mechanism was created to replace the European Financial Stability Facility and the European Financial Stabilization Mechanism. This offered bank recapitalization packages directly to the financial sector, rather than doing so via national treasuries as in the past with existing EU funding programs. In parallel, a Single Supervisory Mechanism was established for the oversight of credit institutions.

4.   Conclusions

4.1   Huge risk misjudgment

This paper aimed to find out why vast masses of individuals and institutions risk their money in ventures that turn out to be a complete fiasco and to explore how to prevent this from happening again in the future. In the three cases analyzed – Argentina's 2001 crisis, the US subprime crisis and the Euro debt crisis – a common feature was the huge misjudgments by investors of the risks really involved. However, in at least two of these three cases, this misjudgment was induced by important actors in the financial world. In the case of Argentina, by the IMF backing of the Convertibility program; in the case of subprime mortgages, by the rating agencies' ratings. The European debt case is a bit more complicated.

In the case of the euro-zone, there was a general assumption that the common currency automatically meant a common level of risk. Having gotten rid of exchange rate risk, investors seemed to assume that sovereign default risks were negligible or in the case national situations worsened, governments would be bailed out by other countries in the euro-zone in order to forestall a breakup of the euro. In other words, the country-specific bankruptcy risk in Europe was either considered to be almost negligible or Article 125 – that says that no country or EU entity can assume responsibility for a member country's public debt – was not taken into due consideration by investors or expected to have a soft interpretation, thus allowing an *in extremis* bailout of debtor countries.

An example of this underestimation of country-specific bankruptcy risk is the following comment in a 2004 article. Speaking on cross-country differences in yields among euro-zone countries, the authors candidly qualified them as striking "as a sovereign default of any of these countries within 10 years seems far-fetched, given their economic history since World War 2" (Pagano and von Thadden, 2004, 550).

This underestimation of default risks has its roots in two huge mistakes. First, the very common one of considering government bonds as almost risk-free assets. As Reinhart and Rogoff (2009) have extensively shown, throughout history rich and poor countries alike have often defaulted on their public debts. Therefore, historical evidence does not support that curiously extended belief.

Second, there was a mistake related to the creation of the euro-zone and its impact on default risk. With a national currency, a government facing a public debt crisis can turn to the central bank and order it to print money and buy up debt. A sovereign default can be avoided at the price of high inflation. In the euro-zone, national governments had transferred monetary sovereignty to the ECB. Therefore, this avenue was closed. The implication is that in a monetary union the probability of a government default is *higher* not smaller than for an isolated individual country government.

Although in the European case the role of rating agencies has been mainly emphasized in connection with the downgrading of European public debt after the crisis burst[30], the fact is that – as in the case of the US financial crisis – the rating agencies long gave overly generous ratings to assets that finally proved to be highly risky and – in the case of Greece – only downgraded them *after* the market had done so. For instance, in Table 1 the mid-2006 S&P's ratings are shown. Ireland's government debt – a country where the banking crisis outburst in 2007 – was still rated as an AAA, with Greece an A and Portugal an AA-.

Table 4.1. Mid-2006 S&P rating

Austria	AAA
Belgium	AA+
Finland	AAA
France	AAA
Greece	A
Ireland	AAA
Italy	AA-
Netherlands	AAA
Portugal	AA-
Spain	AAA

Source:Manganelli and Wolswijk (2007)

Therefore, a key issue for the future is how to protect investors from risk misjudgment.

4.2. The role of the IMF

As stated in the chapter devoted to the analysis of the Argentine case, the IMF played a key role in restoring confidence in Argentina's payment capacity through capital markets. In fact, the misjudgment by the IMF on the sustainability of the Convertibility regime played a key role in reopening Argentina's access to capital markets. The IMF erred in its assessment of the Argentine economy by underestimating the vulnerabilities of the Currency Board regime. Although it was initially reluctant to support the Convertibility regime, which was against the IMF's traditional recipe of a free floating exchange rate, it not only endorsed it but later on even advised other countries – mainly eastern European countries – to adopt it.

The continuous support by the IMF of the Argentine program, even after the Tequila crisis showed the high sensitivity of the Argentine economy to external flows, allowed the government to pile up huge debt, long after it was evident that the Currency Board regime was unsustainable. That support can only be explained by a combination of political and ideological reasons: Argentina had become a star country that was following most of the policies recommended by the Washington Consensus.[31] It was considered that its free markets, deregulation and privatization policies deserved the IMF's support in spite of the inconsistencies in the economic program.

---

[30] An exception may be Iceland where high ratings provided by the rating agencies seem to have played an important role in attracting funds to its bank system. Apparently, these ratings played for investors the role that belonging to the euro-zone provided to its member countries.
[31] "The IMF yielded to external political and market pressures to continue providing its support, despite serious concerns over fiscal and external sustainability" (IMF, 2003, 72).

Because of the weight that political and ideological arguments have in the IMF's decisions as the Argentine case certifies, it is not a reliable source on which investors can be confident. This underlines the need for an independent source of assessment not subject to political or ideological influences. Unfortunately, the next candidate – credit rating agencies – delivered similar or worst results than the IMF.

4.3   The role of credit agencies' ratings

Investors depend on credit ratings to determine the creditworthiness of the assets in which they invest. In the case of institutional investors, it may be argued that, as highly sophisticated investors, they have the capacity to produce their own internal risk analysis. If so, the rating agency's rating would only be used to corroborate the conclusions of their own studies. However, as Keynes (2008) suggests, even professional investment managers have a strong incentive to follow the herd because "it is better to fail conventionally than to succeed unconventionally" (p. 141).

However, there is another reason why it is hard to overstate the importance of the role of credit rating agencies and their ratings: since the mid-1970s, statutes and regulations in the US have increasingly come to depend explicitly on credit agencies' ratings. Therefore, they became regulatory licensors. It was then that rating agencies stopped selling ratings to investors and began charging companies that issue the debt they rate. Regulatory dependence on ratings created higher demand for ratings. However, in several cases their ratings proved spectacularly inaccurate. Prominent examples include California's Orange County and Enron Corp., both of which received high credit ratings until just before they filed for bankruptcy protection. Finally, they certified in large scale as almost risk-free securities assets that were actually highly risky, as the events after 2007 overwhelmingly showed.

In Europe, following the so-called "Basel II recommendations," adopted in 2005, the Capital Requirements Directive introduced a new capital requirements framework for banks and investment firms. The use of credit assessments by External Credit Assessment Institutions was considered to be essential for the determination of risk weights. In essence, it forced European banks and even the ECB itself to rely on the standardized assessments of credit risk provided by credit rating agencies. The new rules on the regulation of credit rating agencies passed by the European Parliament in 2009 restrict banks to use the ratings only for regulatory purposes.[32]

The fact that rating agencies are paid by the issuer raises a conflict of interest. One alternative scheme is investor-paid rating agencies. However, it has been argued that they may also be subject to potential pressure from clients to slide ratings one way or another. Anyway, the experience provided by the US and European crises proves that to rely – as has been argued – only on the self-disciplining role played by reputation makes little sense.

It seems clear that the issuer-paid model does not offer any guarantee to investors. Incentives should be better aligned. A credible threat of civil liability would undoubtedly force credit rating agencies to be more vigilant in guarding against negligent, reckless and fraudulent practices (Partnoy, 2009, 14). Credit ratings should only be part of the mosaic of information considered

---

[32]In May 2011, the European Securities and Markets Authority was assigned the registration and supervision of credit rating agencies in the EU.

to be a part of the investment process. For this purpose, more competition in the industry and the development of new tools to evaluate credit risk seem to be absolutely necessary.

The Dodd-Frank Act passed in July 2010 mandates the SEC to remove ratings requirements for many credit products within a few years. Kurlat and Veldkamp (2011, 3) analyze its effects and conclude that the repeal of ratings mandates will have no effect on the amount of information available about the average security. It would simply transfer the cost of providing the information from the asset issuers to investors.

4.4   Why investors often make the wrong choice?

Besides the misjudgment of risks by institutional actors such as the IMF or credit rating agencies, an additional issue is why investors are frequently attracted by riskier assets. It seems that, as there is "money illusion," there is also "profit illusion" that is, profit is considered without taking into consideration the level of risk involved. Therefore, important portions of capital are usually invested in high-yield high-risk sectors such as the stock market, real estate or assets of dubious quality from tulip bulb contracts to subprime mortgages to Argentine or Greek bonds.

According to prospect theory, as proposed by Kahneman and Tversky, decision makers can become less risk-averse and even risk seeking if they find that they are operating below target or aspiration levels. Laughhunn et al. (1980) examine the behavior of 224 managers from the US, Canada and Europe and find that the majority are risk seekers when faced with below-target outcomes. Strikingly, this picture coincides with the type of behavior described 150 years ago by Marx (2007, 294) according to which the fall in the rate of profit pushes capital "into adventurous channels, speculation, fraudulent credit, fraudulent stocks, crises."

Such behavior also agrees with Minsky's description of investor behavior: "over a protracted period of good times, capitalist economies tend to move from a financial structure dominated by hedge finance units to a structure in which there is large weight to units engaged in speculative and Ponzi finance" (1992, 8).

According to Schumpeter (1961), the primary waves of prosperity initiated by entrepreneurial ventures that implement technological innovations inevitably become overridden by larger secondary waves of speculative prosperity. In his words: "Many things float on this 'secondary wave', without any new or direct impulse from the real driving force, and speculative anticipation in the end acquires a causal significance" (Schumpeter, 1961, 226). Financial crises result from the elimination of speculative ventures and positions but, unfortunately, also of otherwise sound firms that are denied liquidity by now overly cautious bankers. Schumpeter (1950) maintains that 'reckless banking' and financial speculation should be separated from the 'creative destruction' process of innovation by means of "rational as distinguished from vindictive regulation by public authority" (p. 91).

Following Schumpeter's terminology, in the 'primary wave' banks create credit to finance entrepreneurial ventures that introduce new products or processes that increase productivity. However, sooner or later banks find that investment opportunities are running scarce, while savings go on flowing into their vaults. Then, the time for 'financial innovation' comes. One example of financial innovation has been structured finance: in the US, banks neatly packaged multi-trillion dollar dubious mortgages as 'safe' securities and sold them to investors eager to get high yields. Another example of 'secondary wave' financial speculation

and 'reckless' banking was the sale of Argentine bonds by Italian banks to half a million naïve Italian retirees in the 1990s.

These mechanisms are favored if a veil conceals the real risks those investments involve. Here comes the role that rating agencies played in the US subprime financial crisis assuring that those assets were safer than they really were.

However, financial innovation develops only up to the limits that regulations allow. That is why subprime speculation developed *after* financial deregulation took place in the US and not before. For this reason, the 'rational' regulation advocated by Schumpeter should limit 'reckless' banking and speculative excesses.

Although it is true that financial crises can blow up themselves, the severity and social costs of the downturn may be unbearable. Frightened banks severely tighten credit to firms, and this may mean the massive destruction of enterprises and jobs that otherwise would have survived. Alarmed depositors run to withdraw their money from banks, worsening the crisis. Therefore, public authorities should intervene through regulation to avoid that, in Keynes' words, "the capital development of a country becomes a by-product of the activities of a casino" (Keynes, 2008, 142). However, if this is not enough to avoid a financial crisis, government intervention is also necessary to minimize the damage once the crisis blows up. It is always better a soft than a crash landing.

Some may argue that it would be better to let market forces deal with the financial crisis because government intervention creates a moral hazard problem. This was the reasoning behind the denial of a bailout for Lehman Brothers. However, this case precisely showed that one thing is to talk about moral hazard in theory and quite another one to put the idea into practice. After Lehman Brothers' failure, the Fed and the Treasury had to aggressively step in to stop a colossal bank run and rescue the financial system. The argument that troubled banks should not be saved because this would eliminate market participants' incentives to monitor and self-regulate banks' risk behavior proved to be impractical. Given the negative externalities of bank failures due to systemic effects, the social costs of a bankruptcy – particularly in the case of large financial institutions – largely exceed private costs. This puts the onus on regulation in order to minimize the space for moral hazard.

As Keynes (2008, 143) suggests, public access to financial markets should be like access to casinos, "inaccessible and expensive." That is why he argued that the "introduction of a substantial Government transfer tax on all transactions might prove the most serviceable reform available, with a view to mitigating the dominance of speculation over enterprise in the United States" (Ibid.). His idea was that throwing grains of sand into the gears of financial markets might deter financial speculation. However, taxing financial transactions may be a necessary but not sufficient condition for that.

Financial activity as a whole is a public good: systemic risks to financial institutions are risks for the economy as a whole. That is the basic case for the regulation of financial activity. Let us have a look at some of the issues at stake.

4.5   Some issues at stake in financial regulation[33]

There is no doubt that ¨financial regulation is a complex thicket of highly technical policy challenges¨ and that ¨the devil is generally in the details¨ (Véron 2012, 8).

The first issue to be considered is that any regulation means a restriction on the expected rate of return by lowering the level of risk investors or banks are allowed to take. However, this does not necessarily mean a lower *ex post* average rate of return; it only means that riskier bets are excluded or restricted, precisely those that may result in huge losses. Regulation should restrict the types of financial products that financial institutions can offer to the public. It should also include the conditions financial guarantees must meet. The higher transparency of the financial guarantee insurance sector is highly desirable, especially because the assessment of a financial guarantor is further complicated by the presence of an important element of circularity: the values of financial guarantors depend on the values of the securities that they have backed and, in turn, the values of these assets depend on the financial health of the financial guarantor (Schich, 2008, 110).

As stated above, financial institutions have a perverse incentive to take excessive risks. In fact, it is unwise to play safely while everyone else gambles; that is why banks maximize their correlations in order to fail when all other banks are failing, betting that a bailout will take place when a large number of banks are in distress.[34]

Thus, special attention should be placed on those risks capable of damaging the financial system as a whole. This goes beyond the traditional regulatory approach whose primary focus is the safety and soundness of individual institutions and markets in isolation. Systemic significance is not only related to the size of the firm itself but also to its interconnectedness with the rest of the economy. For this purpose, a systemic tax fee – as suggested in Acharya et al. (2009, 284) – for all financial institutions based on their contributions to systemic risk may be a useful tool. This tax would either dissuade financial institutions from those behaviors that increase systemic risk or make them contribute to a fund to be used in the case of a systemic calamity. As in environmental economics, those who pollute must pay for the cost of the clean-up. It is a matter of efficiency and equity.

Milne (2013, 20) argues that "macroprudential tools should be used within a strict rule based framework, in which the impact on the cost and availability of credit can be readily predicted." In this respect, he proposes using "cap and trade" for controlling aggregate systemic liquidity risk instead of the regulation of individual institutions and markets. For the implementation of "cap and trade," a central register of financial assets and liabilities should be established. The systemic risk regulator would determine periodically an amount as the upper limit on the short-term liabilities of financial intermediaries and licenses for this amount would then be distributed to financial institutions. All short-term liabilities used to finance financial investments, both loans and securities, would be subject to licensing control including any offshore funding (Ibid., 5). Exchange between institutions (the trade of licenses) would be allowed to determine the most efficient allocation between institutions. Milne argues that control over the stock of licenses would limit the amount of maturity mismatch in the entire

---

[33] For the dynamics of financial reform as they have unfolded since the start of the crisis see Véron (2012).
[34] Farhi and Tirole (2009, 22) make explicit under what assumptions this is the optimal behavior for banks.

financial system by preventing a rapid increase in the ratio of short-term liabilities to nominal GDP.

4.6   The case of public debt

While regulation can help reduce the level of investors' exposure to risk in the case of private assets, a different issue arises when public debt is involved. How can we minimize the investors' risk of being the victims of a sovereign debt default?

A key issue is transparency in public accounts.[35] However, transparency is not just an issue of making public large quantities of raw data. They must be accessible, relevant and easy for all to understand. Otherwise the public cannot use them to make comparisons and exercise choice. Therefore, the first step is to define the key indicators that allow having a clear idea of fiscal sustainability and a crystal clear way to present them together with a strict schedule for that. For this purpose, the key indicators should also include relevant quasi-fiscal activities conducted outside the general government as well as commitments and contingent liabilities. Pressures to engage in nontransparent practices usually appear during periods of fiscal stress. Therefore, once a schedule has been established, its lack of fulfillment or the delay in reporting on some indicators may be in themselves a signal of fiscal difficulties. If the difficulties are not too serious, the government would probably prefer to air them instead of alarming the financial markets.

An important instrument for ensuring transparency in government operations is an independent review agency responsible for conducting performance audits and studies of selected fiscal issues (Kopits, 2000). To be effective, such an agency must be endowed with wide investigative and reporting authority over government operations.

Finally, as the recent experiences of Iceland, Ireland and Spain illustrate, banking crises may be a cause of sovereign debt crises. Therefore, the health of the banking system is also a critical issue in assessing a country's public debt. Thus, improvement in financial regulation and prudential supervision are not only important for the financial system itself but may also be an important contribution to lower the risk of sovereign debt default.

4.7   A digression: what triggers a financial crisis outburst?

This paper has centered on finding out the reasons why investors risk their money in ventures that turn out to be complete failures and exploring how to avoid this from happening again in the future. That is why very little has been said on the factors that make crises suddenly blow out. However, let me devote a few lines on the research agenda on this subject.

It is difficult to identify the exact factors that determine a crisis outburst. Usually, tensions accumulate during a more or less protracted period until, suddenly, the crisis bursts out. In the case of Argentina, as well as in the cases of Greece, Italy and Portugal, the public debt to GDP ratio was in the dangerous zone several years before the crisis detonated. The outburst seems to have been precipitated by a high share of short-term external debt, which would have required huge doses of creditors' willingness to roll it over. Public debt short maturity structures were also present in the Spanish case.

---

[35]Greece manipulated data to become a member of the euro-zone and concealed the real amount of its public deficit for years until 2009.

The weakening of the government's political power was present in the case of Argentina, where the governing coalition was undergoing a political crisis since the resignation of the vice-president in October 2000, Italy under Berlusconi, Portugal under the minority Socialist government of José Sócrates and Spain under Rodriguez Zapatero. In the presence of a weakening political power, financial markets become suspicious about the ability of the government to collect the taxes needed to fulfill its sovereign debt.

In the case of banking crises, the detonator has usually been the collapse of or the need to rescue an important financial institution, facts that trigger a bank run. The Lehman Brothers' bankruptcy in the US and the nationalization of Iceland's second largest bank were examples of this. The Greek case has something in common with the outbreak of banking crises: it was the sudden disclosure that the amount of its public debt was much higher than it was known up to that moment that triggered a run against Greek public bonds.

Babecký et al. (2012) find that the key early warning signal of crisis comes from growth in domestic credit to the private sector at the horizon of four years. An increase in government debt, the current account deficit and FDI inflow, or a fall in house prices and share prices, are also indicators of an imminent crisis, according to these authors.

Further research may contribute to understanding the specific factors and mechanisms that trigger the outburst of crises. However, I think that the significant question is why financial turmoil can develop until culminating in a crisis without being averted before. Obviously, the key issues are regulation in the case of financial crises and the transparency of public accounts in the public debt case, as stated above.

References

Abell, J.D. (1990):"Twin Deficits During the 1980's: An Empirical   Investigation", *Journal of Macroeconomics,* 12, 81-96.

Acharya, V. V. , Pedersen, L. H., Philippon, T. and Richardson, M. (2009).Regulating Systemic Risk. In Acharya, V. V. and Richardson, M. eds., *Restoring Financial Stability: How to Repair a Failed System.* New York University, Stern School of Business, 283-304.

Aiginger, K. (2010). The Great Recession vs. the Great Depression: Stylized Facts on Siblings That Were Given Different Foster Parents. *Economics: The Open-Access, Open-Assessment E-Journal. Vol. 4, 2010-18.* http://www.economics-ejournal.org/economics/journalarticles /2010-18

Alesina A, and Tabellini G. 1990. A positive theory of fiscal deficits and government debt. *Rev.Econ. Stud.,*57, 403–14.

Alkswani, M. (2000).The Twin Deficits Phenomenon in Petroleum Economy: Evidence from Saudi Arabia. Paper presented at the seventh annual conference, economic research forum (ERF). http://www.erf.org.eg/CMS/uploads/pdf/1185358196_finance1.pdf

André, C. (2010). A Bird´s Eye View of OECD Housing Markets.OECD Economics Department Working Papers, No. 746, OECD Publishing. http://www.cchpr.landecon.cam.ac.uk/Conference2010/Downloads/Papers/Plenary/Andre%20C%20Paper.pdf

Arellano, M. and Bentolila, S. (2009). La burbuja inmobiliaria: causas y responsables. In *La Crisis de la Economía Española: Lecciones y Propuestas*, FEDEA, 28-31. http://www.crisis09.es/PDF/la-crisis-de-la-economia-espanola.pdf

Anoruo, E. and Ramchander, S. (1998). Current Account and Fiscal Deficits: Evidence from Five Developing Economies of Asia. *Journal of Asian Economics*, 9, 487-501.

Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K. and Vašíček, B. (2012).Leading Indicators of Crisis Incidence.Evidence from Developed Countries.ECB Working Paper Series no. 1486. http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1486.pdf

Bachman, D.D. (1992):"Why is the US Current Account Deficit so Large? Evidence From Vector Autoregressions, *Southern Economic Journal*, 59,232-40.

Beker, V. A. (2008).Trade Liberalization, Inequality and Unemployment: Lessons from the Argentine Case**.** http://dx.doi.org/10.2139/ssrn.1126402

Beker, V. A. (2010). On the economic crisis and the crisis of economics. Economics Discussion Papers 2010-18.Kiel Institute for the World Economy.http://econstor.eu/bitstream/10419/36658/1/630979766.pdf

Beker, V. A. and Escudé, G. (2008).Vida, pasión y muerte de la convertibilidad en Argentina. *Estudios Económicos*, Vol. 25, No. 50**, 1-36.** http://bibliotecadigital.uns.edu.ar/scielo.php?script=sci_arttext&pid=S0425-368X2008000100001&lng=es&nrm=iso

Bernanke, B (2010). Causes of the Recent Financial and Economic Crisis.Testimony before the Financial Crisis Inquiry Commission. http://www.federalreserve.gov/newsevents/testimony/bernanke20100902a.htm

Bianchi, J. (2011). Overborrowing and Systemic Externalities in the Business Cycle. *American Economic Review*,101, December, 3400–3426.

Bolton, P. and Jeanne, O. (2011).Sovereign default risk and bank fragility in financially integrated economies.*IMF Economic Review*, 59, 162–194. http://www.palgrave-journals.com/imfer/journal/v59/n2/abs/imfer20115a.html

Bordo, M. D., Eichengreen, B., Klingebiel, D. and Martinez-Peria, M. S. (2001). Is the Crisis Problem Growing More Severe? *Economic Policy* 16, April, 51–82.

Bordo, M. D. and Murshid, A. P. (2001). Are Financial Crises Becoming Increasingly More Contagious? What Is the Historical Evidence? in Forbes, K. and Claessens, S., eds., *International Financial Contagion: How It Spreads and How It Can be Stopped*, New York, Kluwer Academic Publications, 367–406.

Borio, C. and Disyatat, P. (2011).Global imbalances and the financial crisis: Link or no link? BIS Working Papers No 346.

Bracke, T. and Fidora, M. (2012).The macro-financial factors behind the crisis: Global liquidity glut or global savings glut? *The North American Journal of Economics and Finance.* Volume 23, Issue 2, August, 185–202.

Bugamelli, M., Cristadoro, R. and Zevi, G. (2009), La crisi internazionale e il sistema produttivo italiano: un'analisi   su   dati   a   livello   di   impresa, Occasional Papers 58, Banca d'Italia,Rome. http://www.bancaditalia.it/pubblicazioni/econo/quest_ecofin_2/QF_58

Carlin, W. (1998).The new East German economy: problems of transition, unification and institutional mis-match. University College London. http://www.ucl.ac.uk/~uctpa36/East%20Germany%203%20problems.pdf

Cavallo, D. F. (2004). Clase N° 6. Argentina hasta la crisis brasileña. Harvard University. http://www.cavallo.com.ar/wp-content/uploads/C6.pdf

Cavallo, M. (2005).Government Consumption Expenditures and the Current Account. Federal Reserve Bank of San Francisco.Working Paper 2005-03.
 http://www.frbsf.org/publications/economics/papers/2005/wp05-03bk.pdf

Coval, J., Jurek, J. and Stafford, E.(2009).The Economics of Structured Finance. *Journal of Economic Perspectives*: Vol. 23, No. 1, 3-25,Winter.

Desai, P. (2003). *Financial Crisis, Contagion, and Containment: From Asia to Argentina.* Princeton University Press.

Dewald, W.G. and Ulan, M. (1990): The Twin Deficit Illusion. *Cato Journal*, 10, 689- 707.

Di Quirico, R. (2010). Italy and the Global Economic Crisis. Bulletin of Italian Politics.Vol. 2, No. 2, pp. 3-19. http://www.gla.ac.uk/media/media_191024_en.pdf

Dreger, C. and Kholodilin, K.A. (2013).An Early Warning System to Predict Speculative House Price Bubbles. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 7, 2013-8.http://dx.doi.org/10.5018/economics-ejournal.ja.2013-8

Dugan, J. C. (2010). Statement before the Financial Crisis Inquiry Commission. April. http://fcic-static.law.stanford.edu/cdn_media/fcic-testimony/2010-0408-Dugan.pdf

Eichengreen, B., and Lindert, P. H.(1989,), eds.*The International Debt Crisis in Historical Perspective.* Cambridge: MIT Press.

Eichengreen, B. and Park, B. (2012). *The World Economy after the Global Crisis. A New Economic Order for the 21st Century*. World Scientific Studies in International Economics

Éltető, A.(2011). Portugal and Spain: causes and Effects of the crisis.*Baltic Journal of European Studies*, vol.1 no. 2, September, 34/48. http://www.ies.ee/iesp/No10/bjes_no10.pdf#page=34

Enders, W. and Lee, B.S. (1990):" Current Account and Budget Deficits: Twins or Distant Cousins?".*The Review of Economics and Statistics.*72, 373-81.

Erceg, C. and Guerrieri, L.(2005). Expansionary Fiscal Shocks and the Trade Deficit, *Computing in Economics and Finance*, Society for Computational Economics.

Farhi, E. and Tirole, J. Collective moral hazard, maturity mismatch and systemic bailouts. N.B.E.R. Working Paper 15138. http://www.nber.org/papers/w15138

FCIC (2011).*Financial Crisis Inquiry Commission Report*.Official Government Edition.Washington. http://fcic.law.stanford.edu/report

Farhi, E. and andTirole, J. (2009)**.** Collective MoralHazard, Maturity Mismatch and Systemic Bailouts. Toulouse School of Economics Working Paper. http://www.tse-fr.eu/images/doc/wp/ng/wp_nogroup_52_2009.pdf

Feldstein, M. (1991), ed. *The Risk of Economic Crisis.*University of Chicago Press, Chicago and London.

Feldstein, M. (2012).The Failure of the Euro. *Foreign Affairs*, January-February.

Fratianni, M. and Marchionne, F. (2009).The Role of Banks in the Subprime Financial Crisis. Working Papers 2009-02, Indiana University, Kelley School of Business, Department of Business Economics and Public Policy. http://www.bus.indiana.edu/riharbau/RePEc/iuk/wpaper/bepp2009-02-fratianni-marchionne.pdf

Frenkel, R. (2012). Lessons from a comparative analysis of financial crises. INFER Workshop on The Euro: manage it or leave it! June 22-23. http://www.unich.it/docenti/bagnai/euro/

Georgantopoulos, A. G. and Tsamis, A. (2011).The Macroeconomic Effects of Budget Deficits in Greece: A VAR-VECM Approach. *International Research Journal of Finance and Economics,.*Issue 79, Decermber, pp. 156-166.

Gianviti, F., Krueger, A., Pisani-Ferry, J., Sapir, A. and von Hagen, J. (2010). A European Mechanism for Sovereign Debt Crisis Resolution: A Proposal. Bruegel. http://aei.pitt.edu/15123/1/101109_BP_Debt_resolution_BP_clean_01.pdf

Goldstein, M., G. Kaminsky and C. Reinhart (1999), Assessing Vulnerability. Institute for International Finance, Washington, D.C. http://mpra.ub.uni-muenchen.de/13629/

Gorton, G. B. (2008). The Panic of 2007.Federal Reserve Bank of Kansas City, Jackson Hole Conference, August. http://www.kc.frb.org/publicat/sympos/2008/gorton.08.04.08.pdf

Gorton, G.B. (2009). Slapped by the Invisible Hand: The Panic of 2007.Federal Reserve Bank of Atlanta's 2009 Financial Markets Conference: Financial Innovation and Crisis. http://www.frbatlanta.org/news/CONFEREN/09fmc/gorton.pdf

Governor of the Central Bank of Ireland (2010).The Irish Banking Crisis Regulatory and Financial Stability Policy 2003-2008.
http://www.financialregulator.ie/press-area/pressreleases/Pages/TheIrishBankingCrisis,RegulatoryandFinancialStabilityPolicy2003%E2%80%932008.aspx

Icelandic Parliament.Special Investigation Commission. Report. http://www.rna.is/eldri-nefndir/addragandi-og-orsakir-falls-islensku-bankanna-2008/skyrsla-nefndarinnar/english/

International Monetary Fund (2003). Lessons from the Crisis in Argentina. http://www.imf.org/external/np/pdr/lessons/100803.pdf

International Monetary Fund (2004). Report on the Evaluation of the Role of the IMF in Argentina, 1991–2001. http://www.imf.org/External/NP/ieo/2004/arg/eng/

International Monetary Fund and International Development Association (2004). Debt Sustainability in Low-Income Countries-Proposal for an Operational Framework and Policy Implications. http://siteresources.worldbank.org/INTDEBTDEPT/PublicationsAndReports/20261908/debtSust-complete%20paper.pdf

Islam, M. F. (1998). "Brazil's Twin Deficits: An Empirical Examination." *Atlantic Economic Journal,* no. 2, pp. 121-28.

Kartik, N., Ottaviani, M., and Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory* 134(1), May, pp. 93-116.

Katsimi, M. and Moutos, T. (2010). "EMU and the Greek Crisis," *European Journal of Political Economy,* December, 26(4), 568-576.

Khalid, A.M and Teo, W.G. (1999) "Causality Tests of Budget and Current Account Deficits: Cross-Country Comparisons" *Empirical Economics* 24, 389-402.

Kim.K.H. (1995): On the Long-run Determinants of the US Trade Balance: A Comment. *Journal of Post Keynesian Economics.* 17, 447-55.

Keynes J. M. (2008) *The General Theory of Employment, Interest and Money.* Atlantic, New Delhi.
Kopits, G. (2000). Transparency in Government Operations. Interamerican Development Bank.

Krueger, A., 2002, Crisis Prevention and Resolution: Lessons from Argentina. Paper presented at the conference   "The Argentina Crisis," Cambridge, Massachusetts, National Bureau of Economic Research. http://www.imf.org/external/np/speeches/2002/071702.htm

Krugman, P. R. (2009). *The Return of Depression Economics and the Crisis of 2008.* W. W. Norton, New York.

Kurlat, P. and Veldkamp, L. (2011. De-regulating Markets for Financial Information. http://econ.as.nyu.edu/docs/IO/21425/Veldkamp_10062011.pdf

Kwan, S. H. and Laderman, E. S. 1999. On the Portfolio Effects of Financial Convergence—A Review of the Literature. *Federal Reserve Bank of San Francisco Economic Review.* No. 2, pp. 18-31.
Lane, P.R. (2010). The Irish Crisis. IIIS Discussion Paper No.356. http://www.tcd.ie/iiis/documents/discussion/pdfs/iiisdp356.pdf

Lapavitsas, K. (2012). *Crisis in the Eurozone.* Verso, London.

Laughhunn, D.J., Payne, J.W. and Crum, R. (1980). Managerial Risk Preferences for Below-Target Returns., *Management Science* 26, December, 1238-1249.

Leachman, L. L. and Francis, B. 2002.Twin Deficits: Apparition or Reality? *Applied Economics*, Vol. 34, pp. 1121-32.

Leathers, C.G. and Raines, J. P.(2004). The Schumpeterian role of financial innovations in the New Economy's business cycle. *Cambridge Journal of Economics*, September, 667–681.

Lischinsky, B. (2003). The Puzzle of Argentina's Debt Problem: Virtual Dollar Creation? In *The Crisis That Was Not Prevented: Argentina, the IMF, and Globalisation*, .Ch. 6. FONDAD. http://www.fondad.com/uploaded/Argentina/Fondad-Argentina-Chapter6.pdf

Manganelli, S. and Wolswijk, G. (2007). Market discipline, financial integration and fiscal rules: what drives spreads in the euro area government bond market? Working Paper Series 745, European Central Bank. http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp745.pdf

Marx, K. (2007). *Capital. A Critique of Political Economy*. Vol. III, Part I. Cosimo, inc., New York.

McKinsey Global Institute (2011).Debt and deleveraging: The global credit bubble and its economic consequences (Updated Analysis).
http://www.mckinsey.com/insights/mgi/research/financial_markets/debt_and_deleveraging_the_global_c redit_bubble_update

Milesi-Ferretti, G (2009). Notes on the financial crisis and global financial architecture. In G-20 Workshop on the Global Economy, Causes of the crisis: Key Lessons.Mumbai, India 24-26 May. http://www.bancaditalia.it/studiricerche/convegni/atti/ocean_apart/sessione1/1_Milesi_Ferretti.pdf

Miller, S.M. and Russek, F.S. (1989): "Are the Twin Deficits Really Related?" *Contemporary Policy Issues.* 7, pp. 91-115.

Milne, A. (2013). Register, Cap and Trade: A Proposal for Containing Systemic Liquidy Risk.Economics: The Open-Access, Open-Assessment E-Journal, Vol. 7, 2013-7. http://dx.doi.org/10.5018/economics-ejournal.ja.2013-7

Minsky, H. P.(1992). The Financial Instability Hypothesis. The Jerome Levy Economics Institute of Bard College.Working Paper No. 74. http://www.levyinstitute.org/pubs/wp74.pdf

Miranda, M. S. and Calvo Hornero, A. Evolution of the Spanish and Italian Banking in the Financial World Crisis. http://xivrem.ujaen.es/wp-content/uploads/2012/05/14-R-040M403.pdf

Mukhtar, T., Zakaria, M., and Ahmed, M. (2007). An empirical investigation for the twin deficits in Pakistan. *Journal of Economic Cooperation*, 28,4, 63-80.

Pagano, M. and von Thadden, E. L. (2004).The European Bond Markets under EMU. *Oxford Review of Economic Policy*, Oxford University Press, vol. 20(4), pp.531-554, Winter.

Paniza, U. (2013). Do We Need a Mechanism for Solving Sovereign Debt Crises? A Rule-Based Discussion. Graduate Institute of International and Development Studies Working Paper No: 03/2013. http://repec.graduateinstitute.ch/pdfs/Working_papers/HEIDWP03-2013.pdf

Partnoy, F. (2009).Rethinking Regulation of Credit Rating Agencies: An Institutional Investor Perspective. San Diego Legal Studies Paper No. 09-014. http://ssrn.com/abstract=1430608

Pérez-Caldentey, E. and Vernengo, M. (2012).The Euro Imbalances and Financial Deregulation: A Post-Keynesian Interpretation of the European Debt Crisis. Levy Economics Institute of Bard College Working Paper No. 702. http://www.levyinstitute.org/pubs/wp_702.pdf

Piersanti, G., (2000). Current account dynamics and expected future budget deficits: some international evidence. *Journal of International Money and Finance*, Elsevier, vol. 19(2), 255-271, April.

Pozsar, Z., Adrian, T., Ashcraft, A., and Boesky, H. (2012). Shadow banking. Federal Reserve Bank of New York.Staff Report No. 458. http://www.ny.frb.org/research/staff_reports/sr458.pdf

Rajan, R. (2005). ¨Has Financial Development Made the World Riskier? NBER, Working Paper 11728. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=842464

Reinhart, C.M. and Rogoff, K.S. (2004) Serial Default and the ¨Paradox¨ of Rich to Poor Capital Flows. National Bureau of Economic Research. Working Paper 10296. http://www.nber.org/papers/w10296

Reinhart, C. M., and Kenneth S. Rogoff. K. S. (2008a). Banking Crises: An Equal Opportunity Menace. National Bureau of Economic Research Working Paper 14587. http://www.nber.org/papers/w14587

Reinhart, Carmen M., and Kenneth S. Rogoff. 2008b. This Time is Different: A Panoramic View of Eight Centuries of Financial Crises. NBER Working Paper No. 13882 http://www.nber.org/papers/w13882.pdf

Reinhart, C., M.,Rogoff, K. S. and Savastano, M. A. (2003).Debt Intolerance. *Brookings Papers on Economic Activity,* Vol.1, Spring, 1–74.

Robertson, D. (2011). So That's Operational Risk! (How operational risk in mortgage-backed securities almost destroyed the world's financial markets and what we can do about it). OCC Economics Working Paper 2011-1, March. http://www.occ.gov/publications/publications-by-type/economics-working-papers/2011-2009/wp2011-1.pdf

Roubini, N. (2001). Debt Sustainability: How to Assess Whether a Country is Insolvent. Stern School of Business, New York University. http://people.stern.nyu.edu/nroubini/papers/debtsustainability.pdf

Roubini, N. and Mihm, S. (2010). *Crisis Economics: A Crash Course in the Future of Finance.* Penguin Group

Schich, S. (2008). Challenges Related to Financial Guarantee Insurance. *Financial Market Trends,* OECD, 81-113.

Schneider and Kirchgässner (2009).Financial and world economic crisis: What did economists contribute? *Public Choice.* Springer. Vol. 140(3), September, 319-327.

Schumpeter, J. A. (1950). *Capitalism, Socialism, and Democracy,* 3rd ed. New York, Harper and Row.

Schumpeter, J. A. (1961). *The Theory of Economic Development,* New York, Oxford University Press.

Schwarcz, S. (2011).Securitization and Structured Finance. Duke Law & Economics Center. http://www.masonlec.org/wp-content/uploads/2011/03/SCHWARCZ_Encyclopedia-of-Financial-Globalization-Elsevier.pdf?phpMyAdmin=9xTewdhiqRDcmliZ7d4fdRpssn1

Sturzenegger, F. and Zettlemeyer, J. (2006). *Debt Defaults and Lessons from aDecade of Crises* Cambridge: MIT Press.

Taleb, N. N. (2007). *The Black Swan.* Random House.

Tichy, G. (2011)). Could we have foreseen the financial crisis? In Streissler, E. and Tichy, G., eds., *How to Forecast Economic Developments During and After Crises.* Austrian Academy of Sciencies, Vienna.

Tomz, M. and Wright, M.L. J. (2007) Do Countries Default in "Bad Times"? *Journal of the European Economic Association,* 5(2-3), April-May, pp. 352-60.

Véron, N. (20112). Financial Reform after the Crisis, in Eichengreen and Park, eds., *The World Economy after the Global Crisis. A New Economic Order for the 21st Century.* World Scientific Studies in International Economics. Chapter 2.

Zettelmeyer, J., Trebesch, C. and Gulat, M. (2012). The Greek Debt Exchange: An Autopsy. http://ssrn.com/abstract=2144932

Author contact: victor.beker@ub.edu.ar.

_____

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

Global output growth: wage-led rather than profit-led?

Leon Podkaminer   [Vienna Institute for International Economic Studies, Austria]

Abstract
Contrary to 'conventional wisdom' globalization seems to have been associated with slowdown of global output growth and falling share of capital formation (investment) in global output. Referring to the theory of 'demand-led growth', this Note suggests that the negative global tendencies may have arisen under systematic declines in the shares of wage incomes worldwide experienced over recent decades. Making globalization more 'productive' (and investment-friendly) may require a global rebalancing of interests of labor and business.

JEL codes F63, E25, O41

Keywords globalization, investment share, income distribution, growth

Preliminaries

Aggregate gross capital formation (*investment* henceforth) is the central determinant of economic growth. Of course, in the shorter run investment tends to be quite capricious.[1] It responds, rather unpredictably, to psychological factors (e.g. volatile sentiments of entrepreneurs) and to some possibly harder economic influences (e.g. interest rates) – as well as to the perceived opportunities generated by technical progress. Certainly, it is tempting to go beyond analyses of such short-term – accidental or exogenous – influences and attempt to gain some understanding of the factors possibly responsible for the longer-term dynamics of investment and thus of overall growth.

There has been no shortage of theories and concepts relating to the longer-term trends in investment. For this Note the starting point is the theory linking investment dynamics to the functional distribution of output: that is the proportion in which national output (or income) is divided between wages and profits. (The theory abstracts from the fact that income earned by the self-employed – for example farmers – is neither profit nor wage.)

The theory, formally initiated around 1990[2], assumes that in the longer run private investment is an immutable function of two 'variables': (1) the profit share; (2) the level of production capacity utilization. Each of these two variables, taken separately, is assumed to exert a positive impact on investment. However, the level of capacity utilization is higher when the *wage* share is higher (as the consumption propensity out of wage income is 'naturally' higher than the propensity to consume out of profits). Hence the profit share and the level of capacity utilization are not independent of each other – actually these two variables are 'antagonistic'. Depending on some (fairly simple) analytical considerations, it is possible – at least in theory – to identify one of the two variables in question as eventually dominant in so far as investment impacts are concerned. If a certain arithmetical inequality is satisfied then the profit share is dominant, otherwise it is the capacity utilization. In the former case investment (and overall output) growth responds positively to redistribution of income from wages to profits. In the latter case investment (and

---

[1] John Hicks once remarked that 'investment is a flighty bird…which needs to be controlled'.
[2] See e.g. Bhaduri and Marglin (1990), Setterfield et al. (2002), Bhaduri (2007), Lavoie and Stockhammer (2012).

overall output) growth responds positively to redistribution of income from profits to wages. Not surprisingly, the former case is called an instance of 'profit-led growth', and the latter a 'wage-led growth'.

The ambiguous empirics at the national level

The above-outlined theory does not really allow for the existence of 'external world': hence the abstract economy analyzed does not trade with 'the rest of the world'. Nor is it linked to 'the rest of the world' via e.g. capital (including FDI) and capital-related income flows. The fact that transnational corporations' earned profit comes from activities conducted globally, and their investments also cross the borders, must be ignored. Moreover, in the context of progressing integration of national economies (globalization) growth recorded in some countries has come to depend on the net external demand these countries register – and less on what happens to the domestic demand (be it consumption or investment). In the same vein growth in some other countries could have been divorced from trends in their profit shares or capacity utilization levels as domestic consumption and investment may have been fed by growing foreign indebtedness.

The external impacts listed above have been of growing importance, as amply documented: since the late 1960s growth in separate national economies has been increasingly export-led, or import-fed (as the case might be), in addition to being either wage- or profit-led, while cross-border profit- earning and investment activities have been gaining in importance.[3]

Given the strength of internationalization of national economies worldwide, it is perhaps not quite surprising that attempts to characterize growth in *separate* countries as being led by either domestic wages or domestic profits have not produced unambiguous econometric results (see Lavoie and Stockhammer, 2012 for a recent review). Of course, the weaknesses of the available statistics may have played a role as well. The measurement of the wage (or profit) shares at the national levels is easy only in theory. In practice this measurement may be problematic if only because of the existence of the self-employed or working owners whose incomes are hard to classify. The emergence of the class of managers whose exorbitant incomes (actually rents extracted) are formally counted as wages blurs the data even further. In addition, the practices of big multinationals (and wealthy individuals) to declare their incomes (if at all) in places offering tax privileges (rather than where they are actually generated) may play havoc to the profit/wage share statistics of *separate* national economies. In consequence the national data may suggest the absence of relationships between reported national profit/wage shares, capacity utilization levels and national investment growth - even if such relationships actually exist.

---

[3] In 1990 the worldwide stock of FDI is estimated as amounting to 11 per cent of world output – against 35 per cent in 2010. By 2010 the transnational corporations generated about a quarter of global GDP, while their foreign affiliates generated about one tenth of global GDP and one third of global export (UNCTAD, 2012, pp. 24-25).

Figure 1 Global investment as per cent of World GDP, 1970-2009
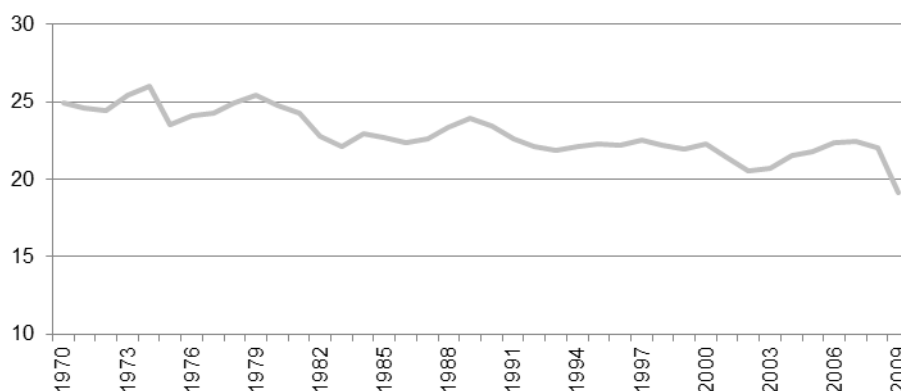
Source: World Development Indicators, World Bank

How about the character of *global* growth?

The national statistics on the shares of wages and profits for separate countries leave much to be desired, as discussed above. But there is little doubt that generally the profit shares have been on the rise – at the expense of the wage shares – in any case since the early 1970s. This fact is amply documented in the recent Report of the International Labor Organization (ILO, 2013, pp. 41-60). According to this source, the average share of labor income in 16 high-income OECD countries fell from 75% of the national income in mid-1970s to about 65% by 2010. The decline in the income share was even more pronounced in many emerging markets (including most New EU Member States[4]) – but also in China and India.

Figure 2 Global output growth rates. 1961-2011
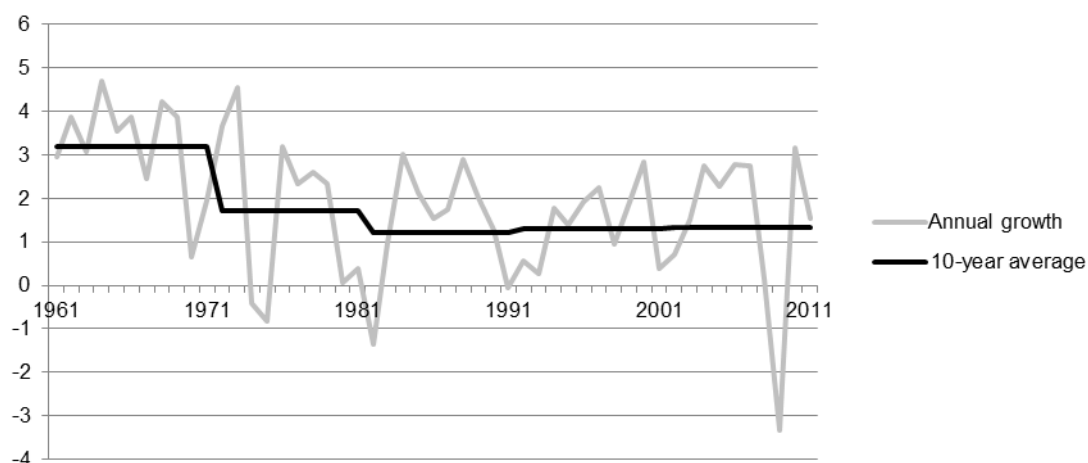
Source: World Development Indicators, World Bank

While the precise statistics on the share of profits (or wages) for the *global* economy remain to be worked out, there is little doubt that globally the share of wages contracts[5] while the share of

---

[4] The strongest decline was registered in Poland where the GDP wage share fell from 68.3% in 1993 to 53.7% by 2011.
[5] Rough calculations conducted at The Vienna Institute for International Economic Studies (based on Eurostat's World Input-Output Database, WIOD) suggest that the global income wage share oscillated between 53% and 53.5% over the years 1995-2000. Thereafter that share declined continually before stabilising at 51-51.5% after 2007.

profits expands. Now, the question is whether or not this 'stylized development' has been associated with investment acceleration at the global level.

As it turns out, investment growth has actually been slowing down relative to global output secularly, since the early 1970s. This fact is reflected in the falling share of global investment in global output (Figure 1).

Concluding remarks

The theory explaining investment (and output) growth by reference to the functional distribution of income (between wages and profits) – but abstracting from complications due to progressing globalization – cannot be reliably tested at the national level. However, the theory could, at least in principle, be tested more reliably at the global level. The stylized fact (rising global share of profits), coupled with a more hard fact (falling global share of investment) suggest that the global economy has been of the wage-led type.

One of the reasons why the global economy's growth has been losing momentum (while at the same time becoming increasingly volatile, as shown in Figure 2) may have been the upset balance between the interests of labor and business – i.e. between wages and profits.

The return to faster, and less volatile, growth globally – and also at the national levels – may require pronounced changes not only as concerns the introduction of regulations restricting the financial sector's disruptive practices. Also, something may have to be done – at national and international levels – to limit the downward drift in wage shares. The proven rule, once obeyed by economic policy making, that wages must move hand-in-hand with labor productivity needs to be resurrected.

References

Bhaduri, A. (2007), 'On the Dynamics of Profit- and Wage-led Growth', *Working Paper* No. 42 of The Vienna Institute for International Economic Studies.

Bhaduri, A. and S. Marglin (1990), 'Unemployment and the real wage: the economic basis for contesting political ideologies', *Cambridge Journal of Economics*, Vol. 14, pp. 375-93.

ILO (2013), *Global Wage Report 2012/13: Wages and Equitable Growth*, Geneva.

Lavoie, M. and E. Stockhammer (2012), 'Wage-led growth: Concept, theories and policies', *ILO Conditions of Work and Employment Series*, No. 41.

Setterfield, M. (ed.) (2002), *The Economics of Demand-Led Growth. Challenging the Supply-side Vision of the Long Run*, Edward Elgar Publishing.

UNCTAD (2012), *World Investment Report 2012*.

Author contact: podkaminer@wiiw.ac.at

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

Striking it richer: the evolution of top incomes in the United States (Updated with 2012 preliminary estimates)

Emmanuel Saez[1]   [University of California, Berkeley, USA]

What's new for recent years?

2009-2012: Uneven recovery from the Great Recession

From 2009 to 2012, average real income per family grew modestly by 6.0% (Table 1). Most of the gains happened in the last year when average incomes grew by 4.6% from 2011 to 2012.

However, the gains were very uneven. Top 1% incomes grew by 31.4% while bottom 99% incomes grew only by 0.4% from 2009 to 2012. Hence, the top 1% captured 95% of the income gains in the first three years of the recovery. From 2009 to 2010, top 1% grew fast and then stagnated from 2010 to 2011. Bottom 99% stagnated both from 2009 to 2010 and from 2010 to 2011. In 2012, top 1% incomes increased sharply by 19.6% while bottom 99% incomes grew only by 1.0%. In sum, top 1% incomes are close to full recovery while bottom 99% incomes have hardly started to recover.

Note that 2012 statistics are based on preliminary projections and will be updated in January 2014 when more complete statistics become available. Note also that part of the surge of top 1% incomes in 2012 could be due to income retiming to take advantage of the lower top tax rates in 2012 relative to 2013 and after.[2] Retiming should be most prevalent for realized capital gains as individuals have great flexibility in the timing of capital gains realizations. However, series for income excluding realized capital gains also show a very sharp increase (Figure 1), suggesting that retiming likely explains only part of the surge in top 1% incomes in 2012. Retiming of income should produce a dip in top reported incomes in 2013. Hence, statistics for 2013 will show how important retiming was in the surge in top incomes from 2011 to 2012.

Overall, these results suggest that the Great Recession has only depressed top income shares temporarily and will not undo any of the dramatic increase in top income shares that

---

[1] University of California, Department of Economics, 530 Evans Hall #3880, Berkeley, CA 94720. This is an updated version of "Striking It Richer: The Evolution of Top Incomes in the United States", *Pathways Magazine*, Stanford Center for the Study of Poverty and Inequality, Winter 2008, 6-7. Much of the discussion in this note is based on previous work joint with Thomas Piketty. All the series described here are available in excel format at http://elsa.berkeley.edu/~saez/TabFig2012prel.xls.

[2] Top ordinary income marginal tax rates increased from 35 to 39.6% and top income tax rates on realized capital gains and dividends increased from 15 to 20% in 2013. In addition, the Affordable Care Act surtax at marginal rate of 3.8% on top capital incomes and 0.9% on top labor incomes was added in 2013 (the surtax is only 0.9% on labor income due to the pre-existing Medicare tax of 2.9% on labor income). The Pease limitation on itemized deductions also increases marginal tax rates by about 1 percentage point in 2013. These higher marginal tax rates affect approximately the top 1%. Hence, among top earners, retiming income from 2013 to 2012 saves about 6.5 percentage points of marginal tax for labor income and about 10 percentage points for capital income. In words, for top 1% earners, shifting an extra $100 of labor income from 2013 to 2012 saves about $6.5 in taxes and shifting an extra $100 of capital income from 2013 to 2012 saves about $10 in taxes.

has taken place since the 1970s. Indeed, the top decile income share in 2012 is equal to 50.4%, the highest ever since 1917 when the series start (Figure 1).

Looking further ahead, based on the US historical record, falls in income concentration due to economic downturns are temporary unless drastic regulation and tax policy changes are implemented and prevent income concentration from bouncing back. Such policy changes took place after the Great Depression during the New Deal and permanently reduced income concentration until the 1970s (Figures 2, 3). In contrast, recent downturns, such as the 2001 recession, lead to only very temporary drops in income concentration (Figures 2, 3).

The policy changes that took place coming out of the Great Recession (financial regulation and top tax rate increase in 2013) are not negligible but they are modest relative to the policy changes that took place coming out of the Great Depression. Therefore, it seems unlikely that US income concentration will fall much in the coming years.

Great Recession 2007-2009

During the Great Recession, from 2007 to 2009, average real income per family declined dramatically by 17.4% (Table 1)[3], the largest two-year drop since the Great Depression. Average real income for the top percentile fell even faster (36.3 percent decline, Table 1), which lead to a decrease in the top percentile income share from 23.5 to 18.1 percent (Figure 2). Average real income for the bottom 99% also fell sharply by 11.6%, also by far the largest two-year decline since the Great Depression. This drop of 11.6% more than erases the 6.8% income gain from 2002 to 2007 for the bottom 99%.

The fall in top decile income share from 2007 to 2009 is actually less than during the 2001 recession from 2000 to 2002, in part because the Great recession has hit bottom 99% incomes much harder than the 2001 recession (Table 1), and in part because upper incomes excluding realized capital gains have resisted relatively well during the Great Recession.

New filing season distributional statistics

Timely distributional statistics are central to enlighten the public policy debate. Distributional statistics used to estimate our series are produced by the Statistics of Income Division of the Internal Revenue Service (http://www.irs.gov/uac/Tax-Stats-2). Those statistics are

---

[3] This decline is much larger than the real official GDP decline of 3.1% from 2007-2009 for several reasons. First, our income measure includes realized capital gains while realized capital gains are not included in GDP. Our average real income measure excluding capital gains decreased by 10.8% (instead of 17.4%). Second, the total number of US families increased by 2.5% from 2007 to 2009 mechanically reducing income growth per family relative to aggregate income growth. Third, nominal GDP decreased by 0.6% while the total market nominal income aggregate we use (when excluding realized capital gains) decreased by 5.5%. This discrepancy is due to several factors: (a) nominal GDP decreased only by 0.4% while nominal National Income (conceptually closer to our measure) decreased by 2%. In net, income items included in National Income but excluded from our income measure grew over the 2007-2009 period. The main items are supplements to wages and salaries (mostly employer provided benefits), rental income of persons (which imputes rents for homeowners), and undistributed profits of corporations (see National Income by Type of Income, Table 1.12, http://www.bea.gov/national/nipaweb/SelectTable.asp).

extremely high quality and final, but come with an almost 2-year lag (statistics for year 2011 incomes have just been published in the summer of year 2013).

In 2012, the Statistics of Income division has started publishing filing season statistics by size of income at http://www.irs.gov/uac/Filing-Season-Statistics These statistics can be used to project the distribution of incomes for the full- year. It is possible to project reliable full-year statistics by the middle of the following year when most of the returns filed before the regular April 15 deadline have been processed by IRS.[4] We have used filing season statistics for 2012 incomes to produce preliminary 2012 estimates. The projection assumes that, in each income bracket, the fraction of tax returns processed by July 2013 for 2012 returns is the same as the fraction of tax returns processed by July 2012 for 2011 returns. Because 2012 statistics are based on a projection, they are preliminary and will be updated in January 2014 when more complete statistics for year 2012 become available.

Text of "striking it richer" updated with 2012 estimates

The recent dramatic rise in income inequality in the United States is well documented. But we know less about which groups are winners and which are losers, or how this may have changed over time. Is most of the income growth being captured by an extremely small income elite? Or is a broader upper middle class profiting? And are capitalists or salaried managers and professionals the main winners? I explore these questions with a uniquely long-term historical view that allows me to place current developments in deeper context than is typically the case.

Efforts at analyzing long-term trends are often hampered by a lack of good data. In the United States, and most other countries, household income surveys virtually did not exist prior to 1960. The only data source consistently available on a long-run basis is tax data. The U.S. government has published detailed statistics on income reported for tax purposes since 1913, when the modern federal income tax started. These statistics report the number of taxpayers and their total income and tax liability for a large number of income brackets. Combining these data with population census data and aggregate income sources, one can estimate the share of total personal income accruing to various upper-income groups, such as the top 10 percent or top 1 percent.

We define income as the sum of all income components reported on tax returns (wages and salaries, pensions received, profits from businesses, capital income such as dividends, interest, or rents, and realized capital gains) before individual income taxes. We exclude government transfers such as Social Security retirement benefits or unemployment compensation benefits from our income definition. Non-taxable fringe benefits such as employer provided health insurance is also excluded from our income definition. Therefore, our income measure is defined as cash market income before individual income taxes.

---

[4] Taxpayers who request a 6-month filing extension *generally* do not file until October 15. Their tax returns are therefore not processed by IRS until the month of November. A substantial fraction of very high income returns use the filing extension. Hence, estimates based on filing season statistics are not exactly equal to final statistics.

Evidence on U.S. top income shares

Figure 1 presents the pre-tax income share of the top decile since 1917 in the United States. In 2012, the top decile includes all families with market income above $114,000. The overall pattern of the top decile share over the century is U-shaped. The share of the top decile is around 45 percent from the mid-1920s to 1940. It declines substantially to just above 32.5 percent in four years during World War II and stays fairly stable around 33 percent until the 1970s. Such an abrupt decline, concentrated exactly during the war years, cannot easily be reconciled with slow technological changes and suggests instead that the shock of the war played a key and lasting role in shaping income concentration in the United States. After decades of stability in the post-war period, the top decile share has increased dramatically over the last twenty-five years and has now regained its pre-war level. Indeed, the top decile share in 2012 is equal to 50.4 percent, a level higher than any other year since 1917 and even surpasses 1928, the peak of stock market bubble in the "roaring" 1920s.

Figure 2 decomposes the top decile into the top percentile (families with income above $394,000 in 2012) and the next 4 percent (families with income between $161,000 and $394,000), and the bottom half of the top decile (families with income between $114,000 and $161,000). Interestingly, most of the fluctuations of the top decile are due to fluctuations within the top percentile. The drop in the next two groups during World War II is far less dramatic, and they recover from the WWII shock relatively quickly. Finally, their shares do not increase much during the recent decades. In contrast, the top percentile has gone through enormous fluctuations along the course of the twentieth century, from about 18 percent before WWI, to a peak to almost 24 percent in the late 1920s, to only about 9 percent during the 1960s-1970s, and back to almost 23.5 percent by 2007. Those at the very top of the income distribution therefore play a central role in the evolution of U.S. inequality over the course of the twentieth century.

The implications of these fluctuations at the very top can also be seen when we examine trends in real income growth per family between the top 1 percent and the bottom 99 percent in recent years as illustrated on Table 1. From 1993 to 2012, for example, average real incomes per family grew by only 17.9% over this 19 year period (implying an annual growth rate of .87%). However, if one excludes the top 1 percent, average real incomes of the bottom 99% grew only by 6.6% from 1993 to 2012 (implying an annual growth rate of .34%). Top 1 percent incomes grew by 86.1% from 1993 to 2012 (implying a 3.3% annual growth rate). This implies that top 1 percent incomes captured just over two-thirds of the overall economic growth of real incomes per family over the period 1993-2012.

The 1993–2012 period encompasses, however, a dramatic shift in how the bottom 99 percent of the income distribution fared. Table 1 next distinguishes between five sub-periods: (1) the 1993–2000 expansion of the Clinton administrations, (2) the 2000-2002 recession, (3) the 2002-2007 expansion of the Bush administrations, (4) the 2007-2009 Great Recession, (5) and 2009-2011, the first two years of recovery. During both expansions, the incomes of the top 1 percent grew extremely quickly by 98.7% and 61.8% respectively. However, while the bottom 99 percent of incomes grew at a solid pace of 20.3% from 1993 to 2000, these incomes grew only 6.8% percent from 2002 to 2007. As a result, in the economic expansion of 2002-2007, the top 1 percent captured two thirds of income growth. Those results may help explain the disconnect between the economic experiences of the public and the solid macroeconomic growth posted by the U.S.

economy from 2002 to 2007. Those results may also help explain why the dramatic growth in top incomes during the Clinton administration did not generate much public outcry while there has been a great level of attention to top incomes in the press and in the public debate since 2005.

During both recessions, the top 1 percent incomes fell sharply, by 30.8% from 2000 to 2002, and by 36.3% from 2007 to 2009. The primary driver of the fall in top incomes during those recessions is the stock market crash which reduces dramatically realized capital gains, and, especially in the 2000-2002 period, the value of executive stock-options. However, bottom 99 percent incomes fell by 11.6% from 2007 to 2009 while they fell only by 6.5 percent from 2000 to 2002. Therefore, the top 1 percent absorbed a larger fraction of losses in the 2000-2002 recession (57%) than in the Great recession (49%). The 11.6 percent fall in bottom 99 percent incomes is the largest fall on record in any two year period since the Great Depression of 1929-1933.

From 2009 to 2012, average real income per family grew modestly by 6.0% (Table 1) but the gains were very uneven. Top 1% incomes grew by 31.4% while bottom 99% incomes grew only by 0.4%. Hence, the top 1% captured 95% of the income gains in the first two years of the recovery. From 2009 to 2010, top 1% grew fast and then stagnated from 2010 to 2011.

Bottom 99% stagnated both from 2009 to 2010 and from 2010 to 2011. Preliminary statistics for year 2012 show that top 1% incomes increased sharply from 2011 to 2012 while bottom 99% incomes grew only modestly.[5]

The top percentile share declined during WWI, recovered during the 1920s boom, and declined again during the great depression and WWII. This very specific timing, together with the fact that very high incomes account for a disproportionate share of the total decline in inequality, strongly suggests that the shocks incurred by capital owners during 1914 to 1945 (depression and wars) played a key role.[6] Indeed, from 1913 and up to the 1970s, very top incomes were mostly composed of capital income (mostly dividend income) and to a smaller extent business income, the wage income share being very modest. Therefore, the large decline of top incomes observed during the 1914-1960 period is predominantly a capital income phenomenon.

Interestingly, the income composition pattern at the very top has changed considerably over the century. The share of wage and salary income has increased sharply from the 1920s to the present, and especially since the 1970s. Therefore, a significant fraction of the surge in top incomes since 1970 is due to an explosion of top wages and salaries. Indeed, estimates based purely on wages and salaries show that the share of total wages and salaries earned by the top 1 percent wage income earners has jumped from 5.1 percent in 1970 to 12.4 percent in 2007.[7]

---

[5] The exact percentage 95% is sensitive to measurement error, especially the growth in the total number of families from 2009 to 2012, estimated from the Current Population Survey. However, the conclusion that most of the gains from economic growth was captured by the top 1% is not in doubt.
[6] The negative effect of the wars on top incomes can be explained in part by the large tax increases enacted to finance the wars. During both wars, the corporate income tax was drastically increased and this reduced mechanically the distributions to stockholders.
[7] Interestingly, this dramatic increase in top wage incomes has not been mitigated by an increase in mobility at the top of the wage distribution. As Wojciech Kopczuk, myself, and Jae Song

Evidence based on the wealth distribution is consistent with those facts. Estimates of wealth concentration, measured by the share of total wealth accruing to top 1 percent wealth holders, constructed by Wojciech Kopczuk and myself from estate tax returns for the 1916-2000 period in the United States show a precipitous decline in the first part of the century with only fairly modest increases in recent decades. The evidence suggests that top incomes earners today are not "rentiers" deriving their incomes from past wealth but rather are "working rich," highly paid employees or new entrepreneurs who have not yet accumulated fortunes comparable to those accumulated during the Gilded Age. Such a pattern might not last for very long. The drastic cuts of the federal tax on large estates could certainly accelerate the path toward the reconstitution of the great wealth concentration that existed in the U.S. economy before the Great Depression.

The labor market has been creating much more inequality over the last thirty years, with the very top earners capturing a large fraction of macroeconomic productivity gains. A number of factors may help explain this increase in inequality, not only underlying technological changes but also the retreat of institutions developed during the New Deal and World War II - such as progressive tax policies, powerful unions, corporate provision of health and retirement benefits, and changing social norms regarding pay inequality. We need to decide as a society whether this increase in income inequality is efficient and acceptable and, if not, what mix of institutional and tax reforms should be developed to counter it.

Table 1. Real income growth by groups

	Average Income Real Growth (1)	Top 1% Incomes (2)	Bottom 99% Incomes (3)	Fraction of total growth (or loss) captured by top (4)
Full period 1993-2012	**17.9%**	**86.1%**	**6.6%**	**68%**
Clinton Expansion 1993-2000	31.5%	98.7%	20.3%	45%
2001 Recession 2000-2002	-11.7%	-30.8%	-6.5%	57%
Bush Expansion 2002-2007	16.1%	61.8%	6.8%	65%
Great Recession 2007-2009	-17.4%	-36.3%	-11.6%	49%
Recovery 2009-2012	6.0%	31.4%	0.4%	95%

Source: Piketty and Saez (2003), series updated to 2012 in August 2013 using IRS preliminary tax statistics for 2012.

Computations based on family market income including realized capital gains (before individual taxes).

have shown in a separate paper, the probability of staying in the top 1 percent wage income group from one year to the next has remained remarkably stable since the 1970s.

Incomes exclude government transfers (such as unemployment insurance and social security) and non-taxable fringe benefits. Incomes are deflated using the Consumer Price Index.

Column (4) reports the fraction of total real family income growth (or loss) captured by the top 1%. For example, from 2002 to 2007, average real family incomes grew by 16.1% but 65% of that growth accrued to the top 1% while only 35% of that growth accrued to the bottom 99% of US families.
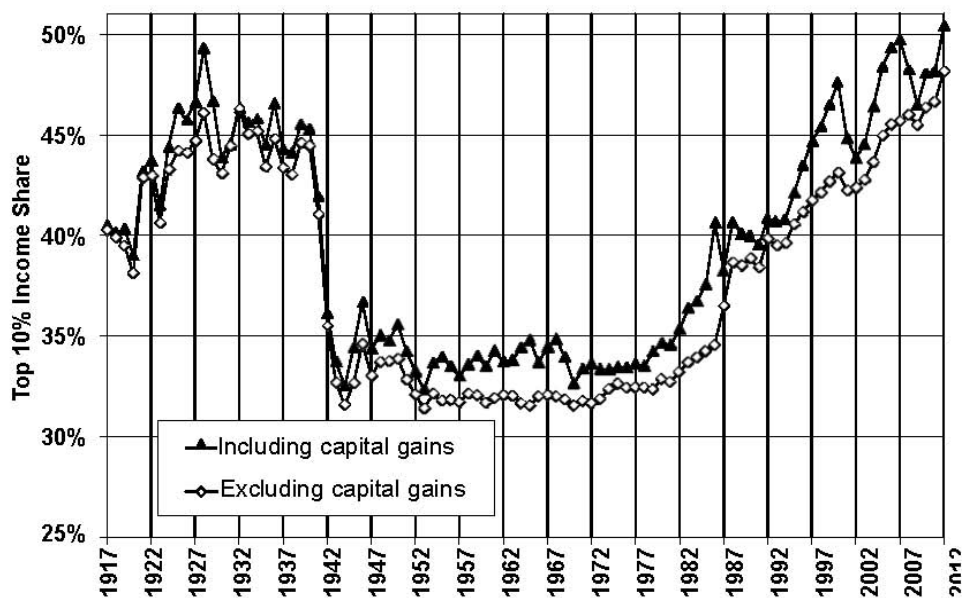
Figure 1: The top decile income share, 1917-2012

Source: Table A1 and Table A3, col. P90-100.
Income is defined as market income (and excludes government transfers).
- In 2012, top decile includes all families with annual income above $114,000.
- 2012 data based on preliminary statistics.
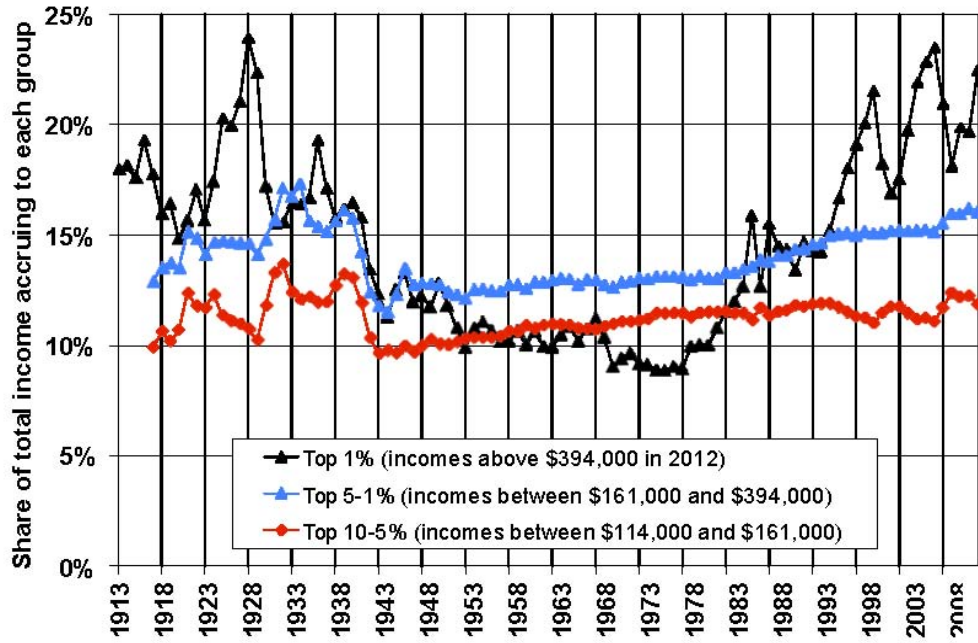
Figure 2: Decomposing the top decile US income share into 3 groups, 1913-2012

Source: Table A3, cols. P90-95, P95-99, P99-100.

Income is defined as market income including capital gains.

- Top 1% denotes the top percentile (families with annual income above $394,000 in 2012)
- Top 5-1% denotes the next 4% (families with annual income between $161,000 and $394,000 in 2012)
- Top 10-5% denotes the next 5% (bottom half of the top decile, families with annual income between $114,000 and $161,000 in 2012).

2012 data based on preliminary statistics.
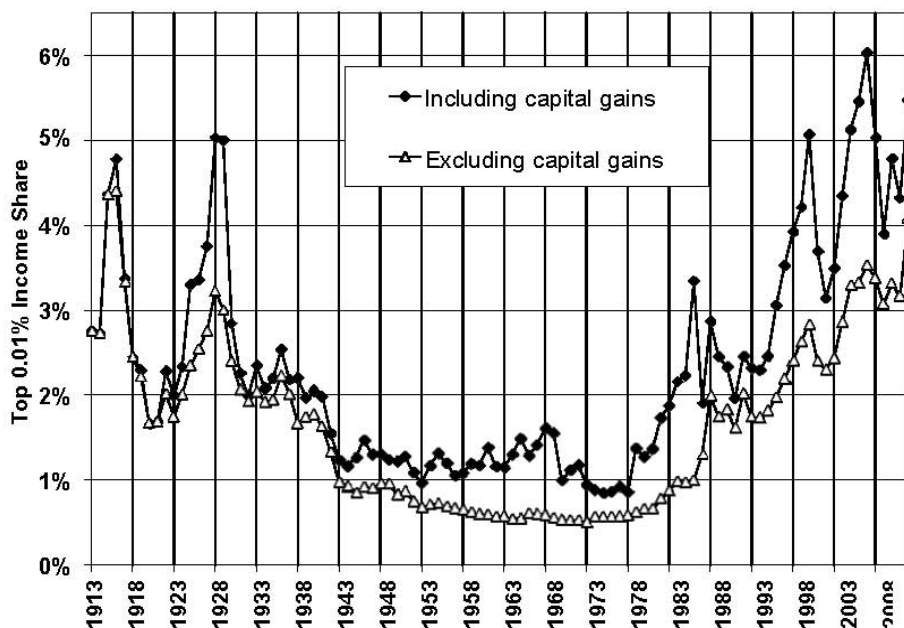
Figure 3: The top 0.01% income share, 1913-2012

Source: Table A1 and Table A3, col. P99.99-100.
Income is defined as market income including (or excluding) capital gains.

- In 2012, top .01% includes the 16,068 top families with annual income above $10,250,000.
- 2012 data based on preliminary statistics

Author contact: saez@econ.berkeley.edu

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

New paradigm economics[1]

Edward Fullbrook   [University of the West of England, UK]

	Old paradigm economics (OPE)	New paradigm economics (NPE)
1[i]	- anti-pluralist (as in classical physics)	- pluralist (as in modern physics)
2[ii]	- prioritizes mathematical deductivism	- recognizes that the ontology of much economic phenomena does not fit the requirements of mathematical deductivism
3[iii]	- beginning with a pure mathematical model, it gives economic entities definitions that make them isomorphic to those mathematical relations. (i.e., upside-down science)	- chooses its math, as in both classical and modern physics, on the basis of its isomorphism to real-world phenomena, including construction of real-world empirical models using real data (i.e., prioritizes the empirical over *apriorism*)
4[iv]	- assumes markets converge toward equilibrium and that therefore theories should be framed around the concept of equilibrium	- recognizes the importance of markets that do not converge toward equilibrium and therefore encourages theory and model development not tied to the equilibrium concept
5[v]	- assumes that when in equilibrium markets have cleared	- does not presume that equilibrium is a market clearing situation
6[vi]	- assumes economic agents have stable preferences and on average behave in a maximizing manner consistent with the neoclassical definition of "rational"	- interested in real-world agent preferences and behavior, "rational" or not, and their macro consequences
7[vii]	- assumes atomistic agents and seeks to explain all meso- and macro-economic phenomena in terms of micro phenomena	- regards agents as social beings, recognizes emergent properties and structures as fundamental to economic reality and thereby the need for a multidimensional ontology
8[viii]	- relies on the ergodic axiom, i.e. reduces uncertainty to risk	- rejects the ergodic axiom, i.e. regards the existence of irreducible uncertainty as a ontological fact that should not be hidden
9[ix]	- treats the planet ("resources") as a subset of the economy	- treats the economy as a subset of the planet and of its biosphere
10[x]	- claims the possibility of a normative-positive distinction in a monist context	- recognizes that the application of any conceptual framework to a real-world economic situation contains a normative or ideological dimension

---

[1] This little piece is indebted to the hundreds of *new-paradigm economists* who have published papers in this journal.

i "Whether you can observe a thing or not depends on the theory which you use. It is theory which decides what can be observed." [Albert Einstein]   Conceptual frameworks and their formalizations attempt to create windows on aspects of the world to the exclusion of others. Pluralism in science generally and in NPE in particular is the belief that x windows, where x > 1, are preferable to only one window. NPE also recognizes that cconceptualizations, including mathematical languages, shape their users' perceptions. For example, economists who use only mathematical languages that do not include positive feedback processes, will see the economy very differently than economists who work with modes of mathematical expression that can accommodate such phenomena.

The NPE recognizes that traditionally fences have stood between economics' various conceptual approaches.   The NPE aims to lower these fences and ultimately, as in physics, to remove them altogether so as to form a common toolbox for understanding economic reality. Such progress entails a movement away from faith-based to empirical-based economics.

ii NPE recognizes that economic reality is characterised ontologically by the property of emergence, whereby there come to exist economic phenomena causally and ontologically irreducible to their components, as in new structures rather than merely new aggregations. This limits the usefulness, as primary methods of approach, of traditional mathematical deductivism and methodological individualism.

iii There are two ways that mathematics can be used relative to an object of inquiry. One is to find or invent a mathematics that fits, i.e. is isomorphic to, the structures and processes of that object. For example, Newton's project of creating classical mechanics was impeded until he invented a mathematics that was isomorphic to the structures he was identifying in the real-world. The other way of applying math or formalism is to make as needed assumptions so as to define elements and combinations of elements in one's empirical realm of enquiry that are isomorphic to a particular mathematics. This of course is upside-down science. It is the math or formalism that determines what structures are going to be attributed to the real-world, rather than real-world structures determining what mathematics, if any, are capable of describing them.   One finds in the original foundation texts of OPE, namely works of Jevons and Walras, the doctrine of upside-down science explicitly and prescriptively spelled out. This has continued to be OPE's dominant approach to the use of mathematics. NPE, on the other hand, rejects upside-down science.

iv   OPE assumes that negative feedback and linear relations always dominate market movements, thereby leading markets toward equilibrium.   NPE recognizes that positive feedback, especially in the new millennium, is often built into market and money supply systems, thereby making the concept of equilibrium at best irrelevant to understanding systems in process. This requires the development of new systems of analysis, informal and formal, with the latter requiring the application of branches of mathematics beyond the scope of OPE.

v When the assumptions of OPE (for example, no institutional factors, no market imperfections, no absence of perfect information, no non-linearities, etc.) are dropped, in other words when real-world situations are considered, then equilibrium or steady-state situations are logically consistent with non-clearing markets, especially unemployment.

vi OPE, with its upside-down methodology, assumes that individual agent behaviour conforms to that particular set of properties necessary for its equilibrium hypothesis to hold mathematically.   NPE recognizes the importance of understanding the impact of numerous categories of economic decisions that violate OPE axioms and which increasingly characterise mainstream economic practice.

vii Physics long ago abandoned doctrinal atomism and the requirement of reductive explanation in favour of an ontology in which fields and forces are also fundamental, mass interchangeable with energy and the properties of particles conditioned by their positional context. Despite its human-realm object of inquiry, OPE's central core remains locked in the metaphysics of 17th century physics. NPE, awoken from "dogmatic slumbers", radically updates economics' ontology by including fundamentally non-atomistic dimensions and non-reductive explanation.   It recognizes the usefulness of sometimes deploying social atomism as a conceptual framework through which to view the economy. But NPE emphasizes the importance, especially in our digital age, of having in our tool box conceptual windows that treat economic agents as social beings, including endogenous preference formations whose interactive structure is integral to the determination of demand.

viii NPE rejects the assumption that there exists a predetermined economic reality that can be fully described by unchanging objective conditional probability functions. NPE favours models set in historical time, thereby generating non-ergodic stochastic processes.   It holds that the OPE methodology is not only ontologically illusionary but also facilitated, by keeping its approach hidden, the Global Financial Collapse.

[ix] NPE regards the economy as dependent on the biosphere and as endangering the composition of its atmosphere, including out-of-control amplifying feedbacks. NPE encourages conceptualizations and analysis that:

1. include the Earth as a "living creature" (Plato),
2. recognize the danger of changing the composition of its atmosphere,
3. recognize the possibility of economic forces doing so, and analyze what is needed economically to prevent the ultimate catastrophe.

[x] NPE opposes attempts to hide the fact that systems of conceptual analysis and applications of their results to real-world economic situations contain a normative or ideological dimension. What one can or cannot see in the world depends on the theoretical lens through which one looks at the world. Therefore different theoretical approaches offer different sets of choices, real or imagined, to be chosen and acted upon by human populations at large. Moreover, unlike in the physical sciences, in the social sciences, economics especially, the conceptual systems used can alter the objects of their enquiries by becoming part of the conceptual and belief systems through which humans conceive of themselves and of others and by which they make choices. These factors impose ethical obligations on the economics profession which the NPE acknowledges, whereas the OPE does not.

Author contact: edward.fullbrook@*btinternet*.com

You may post and read comments on this paper at http://rwer.wordpress.com/2013/09/27/rwer-issue-65/

Articles, comments on and proposals for should be sent to the editor at
pae_news@btinternet.com

Subscriptions to this email journal are free to all members of the World Economics Association (WEA). WEA memberships are free at http://worldeconomicsassociation.org/

Back issues of this journal and material related to the PAE movement are available at www.paecon.net.

To subscribe to this journal go **here** (http://www.feedblitz.com/f/f.fbz?Sub=332386)

To unsubscribe to this journal go to
http://www.feedblitz.com/f/f.fbz?EmailRemove=_NDUzNDkzMHwzMzIzODZ8cGFIX25ld3NAY
nRpbnRIcm5ldC5jb218MTA5OTU4_